

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: May 8, 2002, 12:13:32 ; Search time 28.7 seconds  
(without alignments)  
2293.469 Million cell updates/sec

```

Title:      US-09-647-522-5
Perfect score: 2322
Sequence:   1 MILKHLPMWFIIVLAITSAKH.....SGYIRSMENNPGPGHWSIT 450

```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 segs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

1:  sp.archaea:*
2:  sp.bacteria:*
3:  sp.fungi:*
4:  sp.human:*
5:  sp.invertebrate:*
6:  sp.mammal:*
7:  sp.mhc:*
8:  sp.organelle:*
9:  sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2322	100.0	430	5	Q9GV72	Q9GV72 carboxide ra
2	992.5	42.7	453	5	Q9GNM8	Q9GNM8 carboxide al
3	137	5.9	1222	2	Q9X349	Q9X349 bacillus an
4	117	5.0	833	11	Q62187	Q62187 mus musculi
5	116	5.0	530	2	Q9X0M7	Q9X0M7 thermotoga
6	113.5	4.9	306	2	Q9XR93	Q9XR93 aeromonas p
7	112.5	4.8	510	10	Q9SUX2	Q9SUX2 arabidopsis
8	112.5	4.8	6713	2	Q99U54	Q99U54 staphylococ
9	112	4.8	788	2	Q9F5M7	Q9F5M7 rhizobium m
10	112	4.8	2026	2	Q9JPT4	Q9JPT4 neisseria m
11	111.5	4.8	656	2	Q9WXR0	Q9WXR0 thermotoga
12	111.5	4.8	1336	5	Q94479	Q94479 dictyosteli
13	111	4.8	865	2	Q9XDUI	Q9XDUI clostridium
14	111	4.8	1657	5	Q9BJ59	Q9BJ59 leptomonas
15	110.5	4.8	656	2	Q9WYR0	Q9WYR0 thermotoga
16	110.5	4.8	661	2	Q9X020	Q9X020 thermotoga
17	110	4.7	810	1	P82857	P82857 halobacteri
18	109.5	4.7	874	2	Q06277	Q06277 haemophilus
19	109	4.7	541	2	Q9H0P8	Q9H0P8 pseudomonas

20	109	4.7	612	2	Q9RXX1
21	108.5	4.6	387	2	Q9KXP3
22	107.5	4.6	387	2	Q86266
23	106.5	4.6	1998	2	O50733
24	106	4.6	1369	5	Q9NBM29
25	106	4.6	1975	2	Q9KOS7
26	105.5	4.5	387	2	Q86265
27	105.5	4.5	2504	2	Q85160
28	105	4.5	384	2	Q05492
29	105	4.5	1456	12	Q9IMP2
30	104.5	4.5	333	2	P72126
31	104.5	4.5	406	2	Q9RE67
32	104.5	4.5	532	2	Q9AA46
33	104.5	4.5	739	2	Q44385
34	104.5	4.5	1241	12	Q85444
35	103.5	4.5	349	2	Q9S0U0
36	103.5	4.5	540	2	Q9FCQ4
37	103.5	4.5	565	2	Q9S0T3
38	103.5	4.5	706	9	Q38305
39	103.5	4.5	859	11	Q9JKK5
40	103	4.4	504	2	Q9A8F5
41	103	4.4	638	2	Q36616
42	103	4.4	949	11	Q9DBP9
43	102.5	4.4	460	5	Q17836
44	102.5	4.4	734	5	Q9BIJ2
45	102	4.4	413	2	Q85779
					myxococcus
					Q85779
					Q9RXX1 delnococcus
					Q9KXP3 vibrio chol
					Q86266 pseudomonas
					O50733 borrelia bu
					Q9NBM29 caenorhabdi
					Q9KOS7 neisseria m
					Q86265 pseudomonas
					Q85160 photorhabdu
					Q05492 bacillus ce
					Q9IMP2 portato vtru
					P72126 pseudomonas
					Q9RE67 bacillus ce
					Q9AA46 caulobacter
					Q44385 agrobacteri
					Q85444 diadromus p
					Q9S0U0 escherichia
					Q9FCQ4 pseudomonas
					Q9S0T3 escherichia
					Q38305 lactococcus
					Q9JKK5 mus musculu
					Q9A8F5 caulobacter
					Q36616 vibrio chol
					Q9DBP9 mus musculu
					Q17836 caenorhabdi
					Q9BIJ2 eugruss chl
					Q85779 myxococcus

## ALIGNMENTS

RESULT	1		
09GV72			
ID	09GV72	PRELIMINARY;	PRT; 450 AA.
AC	09GV72;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)		
DE	TOXIN-1.		
GN	CRT-1.		
OS	Carybdea rastoni.		
OC	Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Carybdeidae;		
OC	Carybdea.		
OX	NCBI_Taxid=78582;		
RX	[1]		
RX	SEQUENCE FROM N.A.		
RX	MEDLINE=20422301; PubMed=10964707;		
RA	Nakai H., Takawa K., Nakao M., Ito E., Miyake M., Noda M.,		
RA	Nakajima T.;		
RT	"Novel proteolaceous toxins from the box jellyfish (sea wasp) Carybdea		
RT	rastoni.";		
SR	Biotechnology Abstracts; Res Commun. 27:51587-586(2000)1		
DR	EMBL; AB015878; BAB12728.1;		
SO	SEQUENCE 450 AA; 49392 MW; CD393CF25BEFD2FD CRC64;		

	Query Match	Similarity	Score	DB	Length
Best Local	100.0%	100.0%	2322	5	450
Matches	450	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0

```

Db 181 DSDLSIIIAANVPYKFSNFGLESRIISOGAATSLSDAKRAVDLILYCOLVYMRTEL 240
QY 241 VDLATILYKGNNAHVASAVNANRVNKKELADLDFLHKILIPQALIGAYHHSSETS 300
Db 241 VDLATILYKGNNAHVASAVNANRVNKKELADLDFLHKILIPQALIGAYHHSSETS 300
QY 301 KALINTYKFGVDPVPRPIGNRRYKFTNSYWNYSICSEAYMGNYMFRGCSANRNPIRY 360
Db 301 KALINTYKFGVDPVPRPIGNRRYKFTNSYWNYSICSEAYMGNYMFRGCSANRNPIRY 360
QY 361 SKMSDGYTWMNSDRKKLYITKHDOGWGCTLDEDPDQGMRFIPLRHKYVSSKRW 420
Db 361 SKMSDGYTWMNSDRKKLYITKHDOGWGCTLDEDPDQGMRFIPLRHKYVSSKRW 420
QY 421 NMFYMESSASGYIRSMENPPGQHWST 450
Db 421 NMFYMESSASGYIRSMENPPGQHWST 450

```

```

RESULT 2
O9GNB8 PRELIMINARY: PRT: 463 AA.
AC 09GNB8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE TOXIN-A PRECURSOR.
GN CAT-1.
OS Carybdea alata.
OC Eukaryota; Metazoa; Cnidaria; Cubozoa; Cubomedusae; Carybdeidae;
OC Carybdea.
OX NCBI_TaxID=112899;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20422302; Pubmed=10964708;
RA Nagai H., Takawa K., Nakao M., Sakamoto B., Crow G.L., Nakajima T.;
RT Isolation and characterization of a novel protein toxin from the
RT Hawaiian box jellyfish (sea wasp) Carybdea alata.
RL Biochem. Biophys. Res. Commun. 275:589-594(2000).
DR EMBL: AB036714; BAB12727.1;
KW Signal.
FT SIGNAL 1 18 POTENTIAL.
SQ SEQUENCE 463 AA; 51604 MW; 461253DF6CE9D3C0 CRC64;

```

```

Query Match 42.7%; Score 992.5; DB 5; Length 463;
Best Local Similarity 43.5%; Pred. No. 1,8e-64;
Matches 197; Conservative 95; Mismatches 148; Indels 13; Gaps 7;

```

```

QY 9 LFLVLAITSAGHG-----KSDVNSLITKVEITALKEAS--GSNEAL-LEALEGLKGEIOT 60
Db 11 LFLVLAITSAGHG-----KSDVNSLITKVEITALKEAS--GSNEAL-LEALEGLKGEIOT 60
QY 61 KP-DYGOATKILGSVGSALGLNSGDATKIISGCLDIYAGIATTEGGPYMGIGAVASF 119
Db 71 EPATTTAKVSTIVSGVSGSLKFKSGDPFDVAGSCLDIIASVATTEGPGIAGIYASL 130
QY 120 VSSIIISLFGSSAKNSAAYIDRALSKRDEALORHAAKRPFAESATIQVMKQSNL 179
Db 131 ISSIIISLFGSSAKNSAAYIDRALSKRDEALORHAAKRPFAESATIQVMKQSNL 190
QY 180 TDSDLSTIAANVPYKFSNFGLESRIISOGAATSLSDAKRAVDLILYCOLVYMRTEL 239
Db 191 TEHLDSVRAVYDAFTNMLGYLESTINGSVSTDNENAKRTINFLTLQSLVARETL 250
QY 240 LVDLAILYRK--GNAEHVASAVNANRVNKKELADLDFLHKILIPQALIGAYHHSSETS 297
Db 251 LFGVILLYKAGAGAYDELALSLTSDQNKREATRETVTLFHOHETKYSICSGSYYPIDHS 310
QY 298 ETSKALINTYKFGVDPVPR-PIGNRRYKFTNSYWNYSICSEAYMGNYMFRGCSANRN 356
Db 311 KAAIGILKILTKFGVDPVPR-PIGNRRYKFTNSYWNYSICSEAYMGNYMFRGCSANRN 370

```

```

QY 357 NIKVSKMSDGYTWMNSDRKKLYITKHDOGWGCTLDEDPDQGMRFIPLRHKYVSS 416
Db 371 GIRIKLENGYHTI--TLRSKAMVYTHAOGWGCTLDEDPDQGYTTFIPLTNGFMVST 429
QY 417 KRPWNFMYMESSASGYIRSMENPPGQHWST 449
Db 430 KRPWDYFYMESSAGHYIRSMHYNPPDQGWKI 462

```

```

RESULT 3
O9X349 PRELIMINARY: PRT: 1222 AA.
AC 09X349;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE PXO1-79.
OS Bacillus anthracis.
OC Plasmid virulence plasmid PXO1.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-STERNE;
RA Okinaka R.T., Cloud K., Hamton O., Hoffmaster A., Hill K.K., Kelm P.,
RA Koehler T., Lamke G., Kumano S., Mahillon J., Manter D., Martinez Y.,
RA Riche D.O., Svensson R., Jackson P.J.;
RT "The sequence and organization of PXO1, the large Bacillus anthracis
RT plasmid harboring the Anthrax toxin genes."
RL J. Bacteriol. 0:0-0(1999).
DR EMBL: AF065404; AAD32383.1;
KW Plasmid.
SQ SEQUENCE 1222 AA; 130407 MW; A7C117874D097E5B CRC64;

```

```

Query Match 5.9%; Score 137; DB 2; Length 1222;
Best Local Similarity 23.0%; Pred. No. 0.3;
Matches 107; Conservative 55; Mismatches 173; Indels 130; Gaps 23;

```

```

QY 21 GKRSDV-NSLITKVEITALKEASGSEAL-LEALEGLKGEIOTKPRYGOATKIL--GSVGS 77
Db 310 GVRDDYRNKGVKPRDDEKDEK--TKEDADNPKALCANLGTDTGTSRAARMKAGQVGS 367
QY 78 ALGKLNSGDATKIISGCLDIYAGIATTEGGPYMGIGAVASFVSSILSTFGSSAKNSVA 137
Db 368 ALGKATTTG-----LAGMAA--GAGLSPGAMVAGSOGSALGAAPGAAGRSVA 413
QY 138 AVIDR--ALSKHRDEALORHAAKRPFAESSAFIOVMKQSNLTPSDLS--ITAAVNPV 193
Db 414 AVEGAVNALGKHGSSIKKAGDYYNLNGES-----YPLTDEIDINQDLATDFEN 464
QY 194 YKFSNFGLESRIISOGAATSLSD-----AKRAVDLILYCOLVYMRTEL 243
Db 465 WKADNPNNSAVASLAKOAFNPASDAELAKYAKNSQMSRFTQROKODLQNMKT----- 519
QY 244 ALYKRGNAEHVASAVNANRVNKKELADLDFLHKILIPQALIGAYHHSSETS-- 301
Db 520 ATPY--GNARDLVNAATNA--FQKGYEGDKHDTFMSQLPEN-----MSAEKEKQW 566
QY 302 -----ALNY-----TKYFGV--DVRPIPIGNRRYKFTNSYWNYSICSEAYMGNYM 346
Db 567 NDLNNAKYGFRNHAQAATKAGAMVDAKDKGNN--LFDKSYVKNDAFASDLAALG--- 621
QY 347 FRCGSNVRNPNTIVSKMSGF--TYMENSDR-----RKLYITKHDOGW 388
Db 622 -----KVGKAVGVKGTLESQHONNGOVGALVGASTAFAFKKGYTADHKAAGF- 668
QY 389 WGTLEDPPDQGMRFIPLRHKYVSSKRWPMYFYMESSASGY 433
Db 669 ---MKQFPAD-----MSQEREAAAMKHLDOQYOGF 696

```

RESULT 4  
ID 062187 PRELIMINARY; PRT; 833 AA.  
AC 062187;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE TRANSCRIPTION TERMINATION FACTOR 1 (TRANSCRIPTION FACTOR).  
GN TTF1 OR TTF-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
ON NCBI\_Taxid=10090;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ERLICH ASCITES;  
MEDLINE=95237206; PubMed=7720715;  
RA Evers R., Smid A., Rudloff U., Lottspeich F., Guntt I.;  
RT "Different domains of the murine RNA polymerase I-specific termination factor MTF-1 serve distinct functions in transcription termination.";  
RL EMBL J. 14:1248-1256(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ERLICH ASCITES;  
MEDLINE=95320168; PubMed=7597036;  
RA Evers R., Guntt I.;  
RT "Molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 92:5827-5831(1995).  
DR EMBL: X83974; CA58808.1; -  
DR MGD: MGI:105044; Ttf1.  
DR InterPro: IPR001005; Myb\_DNA\_bind.  
DR Pfam: PF00249; myb\_DNA-binding; 2.  
DR SMART: SM00395; SANT; 2.  
DR PROSITE: PS00090; MYB\_3; 1.  
DR Transcription termination.  
KW SEQUENCE 833 AA; 94534 MW; DEB9C8DC6D3EC6CE CRC64;  
SQ

Query Match 5.0%; Score 117; DB 11; Length 833;  
Best Local Similarity 22.0%; Pred. No. 4.8;  
Matches 100; Conservative 54; Mismatches 127; Indels 174; Gaps 23;

QY 4 KHL-----PMLFYLAITSKHKGRSDVNSLLTYETALAKAGS-SNEALEALEGLK--- 55  
DB 235 KHLQKAPW-----DVGQSQPESISLPPEPLSSDEGKSTEAFAVFCRSLKKNV 286  
QY 56 ---GEIOTKPRDVGQATKILGVSALGKLNCGDATKIISGCLDIYAGIATTFGGPVGNG 112  
DB 287 FNSQLEPIPD-----SLDDESETISERLD-----STHGGAVGAG 321  
QY 113 -----IGAVASFVSSILSLFTSSAKNSVAVAIDR----- 142  
DB 322 ECSTKESHISIKKKKKKKHKSVALATSSDSASVTSKAKN---ALVDSESGAVREED 378  
QY 143 -----ALSKHDEAIQR---HAAGAKRDESSAFIOMVQOQSNLTSDLS 185  
DB 379 VDRPAEAQAQACSTEKHR-EAMORLEPTEHEESNSESASNAARHISEDRESDDVD 437  
QY 186 IIAANPVYKFSNFIGLESRIQGAATSLSDAKRAVDPIILYCOLVYARELLVDLAI 245  
DB 438 LQSA---VQOLREFIPDIDER---AAT---IRRYRDLG- 469  
QY 246 LYRKGAHEVA-----SAVENANRYNKEADTLDFLHKLIPQALIGAVYHPISASET 299  
DB 470 LRREFQAQVAIRFGKFSKAKE-----NKOLEKVVQDL-----SLTGIES 509  
QY 300 SKAILNTYTFYGVDPVPRIGN--RRYKFTNSYWNYSICSEAYMGNYMFGCSNVNPNP 357  
DB 510 ADKLTLTDY---PEKTLITNLRKHAF-----RLHICKGIARPKLV---Y 551  
QY 358 IYVSKMSD-----GEYTMNSDRKRLYITKHHQGW 387

DB 552 YRAKKIPDVNNYKGRYNEEDTKLKAYHSLHGNDW 586  
RESULT 5  
ID 09X0M7 PRELIMINARY; PRT; 530 AA.  
AC 09X0M7;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE METHYL-ACCEPTING CHEMOTAXIS PROTEIN.  
GN TM1143.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogales; Thermotoga.  
ON NCBI\_Taxid=2336;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
RA McDonald L., Uterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";  
RL Nature 399:323-329(1999).  
DR EMBL: AE001771; A036219.1; -  
DR HSSP: P02942; IQ07.  
DR TIGR: TM1143.  
DR InterPro: IPR000122; Chemotaxis\_transducer.  
DR InterPro: IPR000727; T\_SNARE.  
DR Pfam: PF00015; MCPsignal; 1.  
DR PRINTS: PR00260; CHEMTRNSDUCR.  
DR SMART: SM00283; MA; 1.  
DR Complete proteome.  
KW SEQUENCE 530 AA; 57928 MW; C0419A4F46890CE CRC64;  
SQ

Query Match 5.0%; Score 116; DB 2; Length 530;  
Best Local Similarity 19.4%; Pred. No. 2.8;  
Matches 54; Conservative 63; Mismatches 121; Indels 40; Gaps 9;

QY 15 ITSAKHGKRSVDNSLLTKVET---ALKEASGSNEALEALEGKEIOTKPRDVGQATKI 71  
DB 248 ITNQLGISKEMKNTSTRIESIASVQETTAGSEBESISAKTNLADSAQQAAPDOSTOL 307  
QY 72 LGSVGSALGKLNCGDATKIISGCLDIYAGIATTFGGPVGNGIGAVASFVSSILSLFTGSS 131  
DB 308 AKRAGDALKKVI---EVTFRMISNAKDYERYVESF---QKGAELISFVETINAI----- 356  
QY 132 AKNSVAAVIDRALSKRDEAIQRHAAGAKRDESSAFIOMVQOQSNLTSDLSITIANV 191  
DB 357 AEGTNLLANLAALAEAR-----AGEAGRGFAVVADDEIKRLAEESQASQASVRRVYNEI 409  
QY 192 -----PYKFSNFIFIGLESRIQGAATSLSDAK-----RAYVDPIILYCOLVYARET 238  
DB 410 RSIADENAGVSS---EITARVEBGTKLADPADKLNISYGAVERINEMLONIAAIEEOT 466  
QY 239 LLYVD---LAILYRKGAHEVAHAVENANRYNKEADLT 273  
DB 467 AAVDEITTAETAKNAEETITNSVKREVNARLQETISAST 504

RESULT 6  
ID 09R9R9 PRELIMINARY; PRT; 306 AA.  
AC 09R9R9;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)



```

Db 2097 TAKOOLAMSHLTNAOKANLLETOIERGTTVAGVGIQANAGLTNQMNLKOSIASKDAT 2156
QY KIISGCLDIYVGIATTEGPGVMGIGAV-----ASFVSSILSFTG--- 129
Db 2157 KSSEDIYODANADLONAYDAYTNAEGTISATNNPEMNDPTINOKASQVNSAKSALNGDEK 2216
QY 130 -SASKNVAAYIDRALSKHREDAIORHAAKARDEFAESSAFIYVWKQOS----- 177
Db 2217 LAAKQKTKSDIGRLTDLN---AQRTAANAEVQAPPLAATKAKNATSLNTRAMNLT 2273
QY 178 -----NLTDSD-----LSITAAVVPYKFSNFQIOLSRSISQGAATTSLSDAK 220
Db 2274 HALAEKNDTKRSVNTDADQKQOAYDTAVTQAEATITNANGSNANETQVQALNQLNQAK 2333
QY 221 RAYDFILLICQVYVRELLVDLALIRKGAENHVA--SAVENANRVKELAA-DTLDFL 277
Db 2334 NDLEGNKRVAAQ--AKETAKRALASYSNLLNNAOSTAATSOIDNATVTAVTAAQNTANEL 2390
QY 278 HKLIPEQALIGAVHPISASETSKAILNYTKYFQVDPFRIGNRRTSYNNTYSIC 337
Db 2391 N-----TAMQDONGINDONTVAKQVNF-----DADQ--GKDAYTNAVTNAQIL 2435
QY 338 SEA 340
Db 2436 DKA 2438

RESULT 9
Q9F5M7 PRELIMINARY; PRT; 788 AA.
AC Q9F5M7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE METHYL-ACCEPTING CHEMOTAXIS PROTEIN MCPX.
GN MCPX.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RA Muschler P.F.J., Scharf B., Schmitt R.;
RT "Methyl-accepting Chemotaxis protein MCPX of Sinorhizobium meliloti."
RL EMBL: AF312879; AAG34158.1;
DR EMBL: AF312879; AAG34158.1;
DR InterPro: IPR000267; Asparagine_glutamase.
DR InterPro: IPR000122; Chemotaxis_transducer.
DR InterPro: IPR000658; DUF5.
DR InterPro: IPR003660; HAMF.
DR InterPro: IPR003661; His_KINA.
DR Pfam: PF00672; DUF5; 1.
DR Pfam: PF00015; MCPsignal; 1.
DR PRINTS: PR00260; CHEMTRNSDUCR.
DR SMART: SM00304; HAMF; 2.
DR SMART: SM00388; HSKA; 1.
DR SMART: SM00283; MA; 1.
DR PROSITE: PS00144; ASN_GLN_ASE_1; UNKNOWN_1.
SQ SEQUENCE 788 AA; 83753 MW; 3681BC8667DE83F5 CRC64;

Query Match 4.88; Score 112; DB 2; Length 788;
Best Local Similarity 19.58; Pred. No. 10;
Matches 59; Conservative 55; Mismatches 134; Indels 54; Gaps 9;
QY 15 ITSAGKGRSDVNSLTKVETALKEASGNSNALEALLEGKGEIQTGRPDVKGATKILGS 74
Db 518 LITGVNMAAEAKVAASVEVASSDAEGSEVGVKATIAAMGIDQSS---HEVSRIIGV 573
QY 75 VGSALGRLNSGDATKIIISGCLDIYVGIATTEGPGVMGIGAVASVSSILSFTGSSAKN 134
Db 574 I-----DEIAFQTNLLALNAGVEAARAGAGKGFVAVVAGVEVRELAQ--RSANNAK 621

```

```

QY 135 SVAVIDRALSKHREDA-IORHAAKARDEFAESSAFIYVWKQOSNLTDSDLSITIAANPV 193
Db 622 EIKTLINTSAGQVREGVDLVKAGALEKIAEDVOYQIGLIRQISSASSEA-----VGL 676
QY 194 YKFSNFQIOLSRSISOG-----AATSLSDAKRAYDFILLICQVYVRELLVDLAI 245
Db 677 KEINSAVNQMDQVYQONAAWVEETTAASMALENDARLALSALVARQIAPOAAQASAEW 736
QY 246 LYRKGAENHVA SAVENANRVKELAADTLDFLHKLIPEQALIGAVHPISASETSKAILN 305
Db 737 L-RGTERMRRAAPAEER-----PAQADRSAY-----SNSTORVIA 772
QY 306 YT 307
Db 773 KT 774

```

```

RESULT 10
Q9JBJ4 PRELIMINARY; PRT; 2026 AA.
AC Q9JBJ4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PHAB PROTEIN.
GN PHAB.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FAM18;
RX MEDLINE=20187481; PubMed=10722605;
RA Klee S.R., Nassif X., Kusecek B., Werker P., Beretti J.L., Achtman M.,
RA Tinsley C.R.;
RT "Molecular and biological analysis of eight genetic islands that
RT distinguish Neisseria meningitidis from the closely related pathogen
RT Neisseria gonorrhoeae."
RL Infect. Immun. 68:2082-2095(2000).
DR EMBL: AJ391284; CAB72074.1;
SQ SEQUENCE 2026 AA; 210271 MW; D04B9C1EB9C0085A CRC64;

```

```

Query Match 4.88; Score 112; DB 2; Length 2026;
Best Local Similarity 23.58; Pred. No. 44;
Matches 80; Conservative 53; Mismatches 148; Indels 60; Gaps 16;
QY 26 VNSL-----LTKVETALKEASGNSNALEALLEG-LKGEIQTGRPDVKGATKILGSVGSALG 80
Db 1639 VNSFQGEASAKIKTTFSDDYVAKQFA-HALAGCVSGLVQGCKD-----GAIQAVVG 1689
QY 81 K-----LNSGDATKIIISGCLDIYVGIATTEGPGVMGIGAV-ASFVSSILS 125
Db 1690 EIVAEESTLGGNRPATLDAEKHKYISKIIAGSVAALNGDVTMAANAAYVNNMLN 1749
QY 126 L-FTGSSAKNSVAAVIDR-ALSKHREDAIORHAAKARDEFAESSAFIYVWKQOSNLTDSD 183
Db 1750 FDSPTPAKHKQPKPDKTALKEIKIQSIMPAAAGATMNPQDKDAIIVISNIRMGITG-- 1807
QY 184 LSITAAVVPYKFSNFQIOLSRSISQGAATTSLSDAKRAVDLILLYCQVYVRELLVDL 243
Db 1808 -PIVTSYGYVA-AGWAPPLGTAGKKAISTCMNPSGCTVAVTQAAEAGAGIATGATV 1865
QY 244 AIIYRKQNA-EHVASAVENANRVKELAADTLDFLHKLIPEQALIGAVHPIS-ASETSK 301
Db 1866 -----GNANAEAPGALSKAKKAAQALPTQYVREKLDGLDESKNIGAVNFRINIANSTTR 1919
QY 302 AIIINYT--KTFQVP-----DVPRPIGNRRYFTNS 329
Db 1920 -----YTPMROTGPVSAGFEHVLGHPHRIANRSVFTIS 1956

```

```

RESULT 11
ID Q9WXXNO PRELIMINARY; PRT: 656 AA.
AC Q9WXXNO;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE METHYL-ACCEPTING CHEMOTAXIS PROTEIN.
GN TM0023.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwyn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 393:323-329 (1999).
DR EMBL: AE001690; AAD35117.1; -.
DR HSSP: P02942; 1007.
DR TIGR: TM0023; -.
DR InterPro: IPR000122; Chemotaxis_transducer.
DR InterPro: IPR000658; DUF5.
DR InterPro: IPR003660; HAMF.
DR Pfam: PF00672; DUF5; 1.
DR Pfam: PF00015; MCPSignal; 1.
DR SMART: SM00304; HAMF; 1.
DR SMART: SM00283; MA; 1.
DR Complete proteome.
SQ SEQUENCE 656 AA; 72042 MW; 134C37B2B5FCD98 CRC64;

```

```

Query Match 4.8%; Score 111.5; DB 2; Length 656;
Best Local Similarity 20.4%; Pred. No. 8.3; Mismatches 107; Indels 95; Gaps 10;
Matches 63; Conservative 44;

```

```

QY 24 SDVNSLLTVEATLKAAGSNEALALEGL--KGEIQTKPRDVGQATKILGSGSALGK 81
DB 393 NNMSSALTEVTSGEVEAASQINISKITODLTERSEAVTKAREG--TEVEAVGVINK 420
QY 82 LNSG-----DAKIIISGCLDIYAGTA-----TTREGPGMGIGAY 116
DB 451 LKGSARQRDYRELVDASKTIGEIVDTISSIAEQTNLALNAIEMARAGEGRGFAVV 510
QY 117 ASFVSSILSLFTSSAKNSVAAYIDRALSKHDEAIQRHAGAKRDEASSAFIOVMKQ 176
DB 511 ADEIRKL-----AESORATEDIAKMLSTRATTEHENGSKTEFSEVDEIAVWGE 562
QY 177 SMLTDSLSIIANVPYKFSNFIGOLE--SRISOGAATTSIDAKRAVDFTILYCOLV 233
DB 563 -----VTKRRRELIGRIEELNSMIENMTAATAOGGA----- 544
QY 234 VARETLVDAIIIRKGNAEHVAVENANRVKELAADTLDLHLILPEQALIGAVYHP 293
DB 595 -----AEKVASMDNYTKI-VEGVESLNMESLIEDOT----- 627
QY 294 ISASSETSKA 302
DB 628 ESARVSEA 636

```

```

RESULT 12
ID Q94479 PRELIMINARY; PRT: 1336 AA.
AC Q94479;
DT 01-FEB-1997 (TREMblrel. 02, Created)

```

```

DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ABCA (FRAGMENT).
GN ABCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Loomis W.F.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U66526; AAB06789.1; -.
DR InterPro: IPR001687; ATP_GRP_A.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR003593; AAA.
DR Pfam: PF00005; ABC_tran; 1.
DR SMART: SM00382; AAA; 1.
FT NON_TER 1
FT NON_TER 1336
SQ SEQUENCE 1336 AA; 152486 MW; E44ECF939EA53B40 CRC64;

```

```

Query Match 4.8%; Score 111.5; DB 5; Length 1336;
Best Local Similarity 20.8%; Pred. No. 25;
Matches 66; Conservative 49; Mismatches 112; Indels 91; Gaps 14;

```

```

QY 7 PWLFIV-----LAITSAKH-----GKRDVNSLLTVEATLKAAGSNEALALEBGLG 56
DB 334 PMFVLPSYWGSLKRVRIHPPEFDEEDVRAITRAHDA-----SNRAPL-IIIGLSK 386
QY 57 EIQT--KPRVGQATKIL-----GSVGSALGKLTNGDATKI--ISGCLDIYVGIATTFG 106
DB 387 SYTKLEPRPKTYAVAVYVLSLVEKGIILGLGSGCGCKSTTITMLTGLEPTAGDLVYG 446
QY 107 GPVGMGIGAVASFVSI-----LSLPT-----GSSAKNSVAAYIDRA-L 144
DB 447 HSVISNIAAVRITTSVPOHDILMAEMTAREHLQFSELGIPAOBRESQIQKVLQVRL 506
QY 145 SKHRDEAIQRHAGAKR-----DPAESSAFIOVMKQOOS-- 178
DB 507 SKISNNLISTSYSGMKRRUSVALIACIGDKPIIFMDEPTTGVDSSKRHLIDLVKSIKND 566
QY 179 ---LTDSDL-----SIIANVPYKFSNFIGOLESRISOGAATTSIDAKRAVDFTIL 227
DB 567 VILITSHDHEVIELADKIVIMNEGVMACNGNSLQLSKXGEBYSYIVAKSPESIPAVY 626
QY 228 LICQLVY-----MRETL 240
DB 627 EFVTLISIPCKFMKQSGAL 644

```

```

RESULT 13
ID Q9XD01 PRELIMINARY; PRT: 865 AA.
AC Q9XD01;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE 2', 3'-CUCUIC NUCLEOTIDE 2'-PHOSPHODIESTERASE (FRAGMENT).
GN CPD.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13;
RA Bann S., Ohtani K., Yaguchi H., Cole S.T., Hayashi H., Shimizu T.;
RT "Identification of novel VtR/VtRS-regulated genes in Clostridium
RT perfringens.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB028630; BAA81646.1; -.

```

DR HSSP; P07024; 20SH.  
 DR InterPro: IPR002224; 5\_nucleotidase.  
 DR InterPro: IPR001064; Crystallin.  
 DR InterPro: IPR000934; Ser\_thr\_phosphatase.  
 DR Pfam; P00109; 5\_nucleotidase; 2.  
 DR PROSITE; PS00785; 5\_NUCLEOTIDASE\_1; UNKNOWN\_1.  
 DR PROSITE; PS00786; 5\_NUCLEOTIDASE\_2; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETGAMMA; UNKNOWN\_1.  
 FT NON\_TER 865  
 SQ SEQUENCE 865 AA; 95050 MW; 70EB9AFA9CDF5665 CRC64;

Query Match 4.8%; Score 111; DB 2; Length 865;  
 Best Local Similarity 22.6%; Pred. No. 14;  
 Matches 110; Conservative 61; Mismatches 178; Indels 138; Gaps 27;

QY 14 AITSAKHKRSDVNSLTKVETALKESGS---NEALEALEGLK--ETIQKPDYQ 67  
 DB 280 AITSEPARAKYSEKNTLVE---KDESGTVLDKNADLILSKVGQDPLETKKPPHE 336  
 QY 68 ATKILGSVGSALGKINSGLDIAVAGT-----ATTEGCPYGM----- 111  
 DB 337 AA--INDATAKIGELKGDIAK---PDEVKGIPOSIVEDOGVYDFINEVOLYNSKKFL 389  
 QY 112 --GIGAVASVSSILSLFT-----GSSAKNSVAAY--IDRALSKHDEAIORHAGA 159  
 DB 390 QTKGIDPNVNYWSSALFSPKALKEGPIKADVSNIYKFDNKL---YVKTNGKQL 444  
 QY 160 KRDEASSAFIQVKQOQSNLDSLSI--IAANVPYKESNFIQ--QLESRIQ--GAATTS 215  
 DB 445 KKYMEENSKFENKFK---DGDLTISFDENVRYKYKDMFEGVYELINAKDGERIEN 498  
 QY 216 L---SDAKRAVDLILQLVYMETLLVDAIILYRKNAHVASAVENANRVKELAAD 272  
 DB 499 LKFSKDKGPEVSDVYLYSL-----NDYRNSGLAAG 530  
 QY 273 TLDFLHKLIPQALIGAVYHPISASE--TSKAILNTKRYGVPVPRPIGR---RYKF 326  
 DB 531 IMDSGE---HEKITYDVNDISAIKRLISDIYIN-----VKHGVINRNDGKMKI 577  
 QY 327 TNSIYN-----TYSICSEAVMGNYMREGCSNVRNPNI---VSKMSDGYFTMENS 374  
 DB 578 TGNMNMDEORALAVKLINE---GKIKLPTSNGRTPVKSVTWGEVGSFAAL--PEEKE 633  
 QY 375 RRRKIYIKHQGNGWGLDEPQDGHMRPIPLRHGKYMSSKMPNF-----MTNES 428  
 DB 634 VEIPILTFND---FHGSLKESGNPAAKFV---GELKKVKEKNPPTIIVSGGDMYOGS 686  
 QY 429 SASGYIR 435  
 DB 687 ALSNMLK 693

RESULT 14  
 Q9BJ59 PRELIMINARY; PRT; 1657 AA.  
 AC Q9BJ59;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE RNA POLYMERASE II LARGEST SUBUNIT.  
 GN RPAPII.  
 OS Leptomonas seymouri.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leptomonas.  
 OX NCBI\_TaxID=5684;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21169359; PubMed=11266558;  
 RA Gilling G., Belliofatto V.,  
 RT "Trypanosome spliced leader RNA genes contain the first identified RNA  
 RL Nucleic Acids Res. 29:1556-1564(2001).  
 DR EMBL; AF338253; AAK15346.1; -

SO SEQUENCE 1657 AA; 182912 MW; B9CA22C166AF8AD6A CRC64;

Query Match 4.8%; Score 111; DB 5; Length 1657;  
 Best Local Similarity 24.1%; Pred. No. 38;  
 Matches 95; Conservative 51; Mismatches 126; Indels 122; Gaps 20;

QY 65 VGQATKILGSVGSALGK-----NSGDATKIISGLDIAVAGIATTFGCPYGM----- 111  
 DB 592 VGATTK--GVGAAGSLIHVIFNERSGDEVAKFINGVQRTITTYFNKYCFASFVGVQDTYA 649  
 QY 112 -----GIGAVASVSSILSLFTGSSAKNSVAAYIDRALSKHDEA 151  
 DB 650 DASTLKEMNNVLUHKTROSEVETIGAAAN--NGKLTAKAGMSLQSPEDAVNSALNCKREA 707  
 QY 152 IORHAGAKRDPABSSAFIQVKQOQSNLDSLSI--IAANVPYKESNFIQ--QLESRIQ--GAATTS 207  
 DB 708 AKKALSNVR---TNSKVMIEAGSKGSDLNICQIAV-----FVGQNNVAGSRIP 754  
 QY 208 -----SQAATTSLSDAKRAVDLILQLVYMETLL-----VD 242  
 DB 755 GFRRRLPHMLDDYGETSRGMATRGVEGLQPHF---YFHTWAGREGLIDTAVKTS 811  
 QY 243 LAIILYRK--GNAHVASA---VENANRVKELADTLDFLHKLIPQALIGAVYHPISA 296  
 DB 812 TGYLRKLIRKLEVDYHAAYDCTVRNANOELIQLA-----YGEDGLDGA-----R 855  
 QY 297 SETSKAILNTKRYGVPDPV--RPIGNRRYKFTSNYMTYSICSEAVMGNY--FRGCSN 352  
 DB 856 IEGNQA-----PPHPMANNMIDKRYEY-----NEESSFENNMGSGSYMDPVRSL 903  
 QY 353 VRNPNIRYKMSDGF--VTMENSDDRKLITYTKHQ 385  
 DB 904 LRDPQ--SVSKLOEEDQLVXKDRDMSRLTIIDMEK 936

RESULT 15  
 Q9WYR0 PRELIMINARY; PRT; 656 AA.  
 AC Q9WYR0;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE METHYL-ACCEPTING CHEMOTAXIS PROTEIN.  
 GN TW0429.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton M.G., Fleischmann R.D., Eisen J.A., White O.,  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 RT genome sequence of Thermotoga maritima."  
 RL Nature 399:323-329(1999).  
 DR EMBL; AE001721; AAD3514.1; -  
 DR HSSP; P02942; 1007.  
 DR TIGR; TM0429;  
 DR InterPro: IPR000122; Chemotaxis\_transducer.  
 DR InterPro: IPR000658; DUF5.  
 DR InterPro: IPR003660; HAM.  
 DR Pfam; PF00672; DUF5; 1.  
 DR Pfam; PF00015; MCPSignal; 1.  
 DR SMART; SM00304; HAM; 1.  
 DR SMART; SM00283; HAM; 1.  
 KW Complete proteome.  
 SO SEQUENCE 656 AA; 72217 MW; EB932EC2C3505B143 CRC64;





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 12:13:12 ; Search time 13.31 Seconds  
(without alignments)  
1239.607 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322  
Sequence: 1 MILKHPMFLVLAITSAKH.....SGYIRSENNPQGHMSIT 450

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116.5	5.0	376	1 MDHC_YEAST	P22133 saccharomyc
2	112.5	4.8	964	1 PMPE_CHLTR	O64877 chlamydia t
3	110	4.7	1073	1 ACA9_ARATH	O91u41 arabidopsis
4	108.5	4.7	1023	1 HLVA_ECOLI	P09983 escherichia
5	108.5	4.7	1024	1 HLVA_ECOLI	P08715 escherichia
6	105.5	4.5	312	1 FAS3_RHOFA	P46375 rhodococcus
7	105	4.5	306	1 AER_ECOLI	P50466 escherichia
8	103	4.4	1456	1 RRPO_PVXCP	P22591 potato viru
9	102	4.4	393	1 FLAA_PSEAE	P21184 pseudomonas
10	102	4.4	1456	1 RRPO_PVXHB	O07630 potato viru
11	101.5	4.4	299	1 PYRB_ARCFU	O30130 archaeoglob
12	101	4.3	535	1 HTR1_HAUNI	P33741 halobacteri
13	100	4.3	437	1 YABE_BACSU	P37546 bacillus su
14	100	4.3	444	1 YABE_BACSU	O92798 chlamydia p
15	99	4.3	948	1 PMAS_ARATH	O9s1b3 arabidopsis
16	99	4.3	1935	1 PMAS_CYCA	O90339 cyprinus ca
17	98.5	4.2	379	1 FLAA_VIRCH	O30858 vibrio chol
18	97.5	4.2	487	1 SR54_ENTHI	O15821 entamoeba h
19	97	4.2	825	1 GUN3_BACSA	P19570 bacillus sp
20	97	4.2	1934	1 MYH7_MESAU	P13540 mesocricetu
21	97	4.2	1935	1 MYH7_RAT	P02564 ratulus norv
22	95.5	4.1	3027	1 POLG_PYFV1	O05057 parsnip yel
23	95	4.1	381	1 YBD2_YEAST	P38199 saccharomyc
24	95	4.1	434	1 T147_HUMAN	O60664 homo sapien
25	95	4.1	483	1 K1CL_MOUSE	O64291 mus musculu
26	95	4.1	675	1 ATK8_DEIRA	O9r2p0 deinosococu
27	95	4.1	1120	1 STFR_ECOLI	P76072 escherichia
28	95	4.1	1456	1 RRPO_PVX	P09395 potato viru
29	94.5	4.1	548	1 SR54_AOUAE	O67645 aquifex aeo
30	94.5	4.1	548	1 HLVB_VIRCH	P15492 vibrio chol
31	94.5	4.1	1706	1 CYAA_BORPE	P15318 bordetella
32	94	4.0	374	1 FLAI_BARBA	P35633 bartonella
33	94	4.0	450	1 ADFP_BOVIN	O9tun6 bos taurus

34	94	4.0	557	1 MCP5_ENTAE	P21822 enterobacte
35	94	4.0	1935	1 MYH7_HUMAN	P12883 homo sapien
36	93.5	4.0	427	1 HXSX_HAETN	P44001 haemophilus
37	93.5	4.0	580	1 IPAB_SHITL	P18011 shigella fl
38	93.5	4.0	903	1 MSP1_SCHPO	P87320 schizosacch
39	93.5	4.0	1807	1 TSC2_HUMAN	P49815 homo sapien
40	93	4.0	488	1 FLIC_PSEAE	P72151 pseudomonas
41	93	4.0	845	1 VFPA_RHTSN	P55439 rhizobium s
42	93	4.0	1295	1 BXA_CLOBO	P10845 clostridium
43	92.5	4.0	431	1 K1CQ_HUMAN	O04695 homo sapien
44	92.5	4.0	553	1 MCP2_SALTY	P02941 salmoneilla
45	92.5	4.0	1790	1 LMBI_DROME	P11046 drosophilla

#### ALIGNMENTS

```

RESULT 1
MDHC_YEAST STANDARD: PRT: 376 AA.
ID MDHC_YEAST
AC P22133:
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE MALATE DEHYDROGENASE, CYTOPLASMIC (EC 1.1.1.37).
DN MDH2 OR YOL126C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-16.
RP MEDLINE=91094852; PubMed=1986231;
RX Minard K.I., McAlister-Henn L.;
RA "Isolation, nucleotide sequence analysis, and disruption of the MDH2
RT gene from Saccharomyces cerevisiae: evidence for three isozymes of
RT yeast malate dehydrogenase."
RL Mol. Cell. Biol. 11:370-380(1991).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / FY1679;
RX MEDLINE=97051588; PubMed=8896265;
RA Casamayor A., Khalid H., Balcells L., Aldea M., Casas C., Herrero E.,
RA Ariño J.;
RT "Sequence analysis of a 13.4 kbp fragment from the left arm of
RT chromosome XV reveals a malate dehydrogenase gene, a putative Ser/Thr
RT protein kinase, the ribosomal L25 gene and four new open reading
RT frames."
RL Yeast 12:1013-1020(1996).
[3]
RN MEDLINE=OF 1-34.
RP MEDLINE=87185517; PubMed=3552052;
RX Kopeckí E., Entian K.-D., Lottspeich F., Mecke D.;
RA "Purification procedure and N-terminal amino acid sequence of yeast
RT malate dehydrogenase isoenzymes."
RL Biochim. Biophys. Acta 912:398-403(1987).
-1- FUNCTION: THE ISOENZYME MDH2 MAY FUNCTION PRIMARILY IN THE
GLYOXYLATE CYCLE.
-1- CATALYTIC ACTIVITY: L-MALATE + NAD(+) = OXALACETATE + NADH.
-1- SUBUNIT: HOMODIMER.
-1- SUBCELLULAR LOCATION: CYTOPLASMIC.
-1- INDUCTION: BY ACETATE AS CARBON SOURCE IN THE GROWTH MEDIUM.
-1- IS INACTIVATED BY ADDITION OF GLUCOSE (CARABOLITE INACTIVATION).
-1- MISCELLANEOUS: YEAST CONTAINS AT LEAST 3 MALATE DEHYDROGENASE
ISOENZYMES: A MITOCHONDRIAL (MDH1), A CYTOPLASMIC (MDH2) AND A
PEROXISOMAL (MDH3).
-1- SIMILARITY: BELONGS TO THE LDH FAMILY. MDH SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M62808; AAA34766.1; -  
 DR EMBL: U41293; AAC49466.1; ALT\_INIT.  
 DR EMBL: Z74868; CA99145.1; ALT\_INIT.  
 DR PIR: S12937; DEBYMC.  
 DR PIR: S05770; S05770.  
 DR HSSP: P00346; IMLD.  
 DR SGD: S0005486; MDH2.  
 DR InterPro: IPR001252; MDH.actsite.  
 DR InterPro: IPR001236; 1dh.  
 DR Pfam: PF00056; 1dh.  
 DR PROSITE: PS00068; MDH.1.  
 DR Oxidoreductase: Tricarboxylic acid cycle; NAD.  
 FT INIT\_MET 0  
 FT ACT\_SITE 181  
 FT BINDING 184  
 FT ACT\_SITE 214  
 SQ SEQUENCE 376 AA; 40599 MW; C08C63BF85E457EE CRC64;

Query Match 5.0%; Score 116.5; DB 1; Length 376;  
 Best Local Similarity 20.2%; Pred. No. 0.54;  
 Matches 72; Conservative 65; Mismatches 125; Indels 95; Gaps 17;

QY 13 LAITSAGHGRSDVNSLL-TKVEALKEASGS-----NEALELEGLKEGIOTKPD 63  
 DB 15 IALIGAAGGGQSLKALKAQLOQLKESNRSVTHILALYDNOEALINGYADLSHIPT 74  
 QY 64 RVGQATKILGVSALGKLSGDATKILISG-----CLDIVGIATTEGPGVMG 112  
 DB 75 PIVSSSH--SPAGIENCLNNAISIVIPAGVPRKRGMTRODLEFVNNGIISQDLSI--- 129  
 QY 113 IGAVASVSSILFTGSSAKNSYAVIDRALSKH---RDEAIORHAG-AKRDEFAESSA 168  
 DB 130 --AECCCLSYFVLYVINPNSILVPVVSNTLKNHPQSRSGIERRMGTGKLDIVRAST 187  
 QY 169 FIVGMKQOQSNLTD-----SPLSTIAAN-----VPVKEFSNFIGOL----- 203  
 DB 188 FLKRNIESGLTPRVNSMPDVPYVGGHSGEITIFLFQSQNFSLANDQLKYLIRVOYG 247  
 QY 204 ----ESRISGAATTSLSDA--KRAVDFILYQOLVYVMEFTLLVDLAILYRKGNAEHVA 256  
 DB 248 GDEVYAKKNGKSGATLSMAHAGYKCVQFV-----SLL-----GNIDQIH 288  
 QY 257 SA----VENNRVVKELAADTLDLKLIPEDALIGAVYHPISASESKAILNTKY 309  
 DB 289 GTYYVPLKDNANPFIADGAD-----QLLP--LVDGADYFAIPLTITTKGV-STVDY 336

RESULT 2  
 PMPE\_CHLTR STANDARD; PRT; 964 AA.  
 AC 084877;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROBABLE OUTER MEMBRANE PROTEIN PMPE PRECURSOR (POLYMORPHIC MEMBRANE  
 DE PROTEIN E).  
 GN PMPE OR CT869.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID-813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D/UM-3/CX;  
 RX MEDLINE-99000809; PubMed-9784136;  
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.,  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 PT Chlamydia trachomatis.";

RL Science 282:754-759(1998).  
 CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)  
 CC (POTENTIAL)  
 CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AE001360; AAC68467.1; -  
 DR InterPro: IPR003368; DUF145.  
 DR InterPro: IPR003357; OMP.  
 DR Pfam: PF02415; DUF145.1.  
 DR Pfam: PF02385; OMP.1.  
 KM Outer membrane; Signal; Multigene family; Complete proteome.  
 FT SIGNAL 1  
 FT CHAIN 19 964  
 SQ SEQUENCE 964 AA; 104703 MW; 1B998A7D2E571CE2 CRC64;

Query Match 4.8%; Score 112.5; DB 1; Length 964;  
 Best Local Similarity 21.3%; Pred. No. 3.5;  
 Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

QY 21 GKRSDV---NSLTKVETALKEASGSNEALALEGLKEGIOTKPDVQATKILGVS 76  
 DB 254 GNGNIVFYNNRCFKNVEFASSEAS-----DGAIKVTRLDVDTGNRGRIFFS- 301  
 QY 77 SALGKLSNGATKTIISGCDIVAGIATTPGCPYGMGGAASVSSILFTGSSAKNSV 156  
 DB 302 -----DNITKNVGGAI---YAPVTLVDNGPTFYINNIAKNG 336  
 QY 137 AAVIDRALSKRDEAIORHNA-----AGAKR-----DEFAESSAFIQV 172  
 DB 337 GAIYIDGTSNKSISA-DRAHIIIFNENIVTNTANGSTSANPRRAIIVASSGELL 395  
 QY 173 MKQOS-NLTDSD-LSITANVPYKFSNFTGQLESRSOGAATTSLSDAKRAVD----- 224  
 DB 396 GAGSQNLIFIDPIEVSNAQVSV-SFNKADQYGSVFSGATVNSADFHORNIQTKTPAP 454  
 QY 225 -----FILL--YCOLVVMKEL--LVLD-----AIL--YKGNAEHVAEVENNRVKE 268  
 DB 455 LITLSNGFLIEDHQAQLTVNRFOTGGVSLGNAVLSCTYNGTGD-----SASNASTLKH 510  
 QY 269 LAADTLDLKLIPEDALIGAVYHPISASESKAILNTY---KYGVPDVPRI---G 320  
 DB 511 IGLN---LSSILKSAEI-----PLTWPEPTNNSNNTADTATFSLSDVKLSLIDYG 561  
 QY 321 NRRIKFTNSYNTWNTSICSEAYMGNYMRGCSNVRNPIRVSKKSDGYTTEMNSDRRLYI 380  
 DB 562 NSPYESTIDL--THALSSQ-----PWLSTISEASDNOLOEENIDFSGLNV 602  
 QY 381 TKHD-OG--WGNG-FLDEPQDGHMR-----FIPU--RHGYVM 414  
 DB 603 PHIGWGLWTGNAKTODPPASSAITTDPOKANRFRITLLTWLPAGVYSPKHKRSPLI 662  
 QY 415 SSRKWPMEYMESSASGYIRSWENNGPOGHSIT 450  
 DB 663 ANTLWGMMLLATESLKN---SALFLPSGHPFGIT 694

RESULT 3  
 ACA9\_ARATH STANDARD; PRT; 1073 AA.  
 ID ACA9\_ARATH  
 AC 091041;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE POTENTIAL CALCIUM-TRANSPORTING ATPASE 9, PLASMA MEMBRANE-TYPE

DE (EC 3.8.3.8) (CA2+-ATPase, ISOFORM 9).  
 GN AC9 OR AT3G21180 OR M18.3.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eustroids II; Brassicales; Brassicaceae; Arabidopsis.  
 RN NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV\_COLUMBIA;  
 RX MEDLINE=20277480; PubMed=10819329;  
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
 RT features of the regions of 4,504,864 bp covered by sixty pl and YAC  
 RT clones.";  
 RL DNA Res. 7:131-135(2000).  
 CC -1- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS  
 CC OF ATP COUPLED WITH THE TRANSLLOCATION OF CALCIUM FROM THE CYTOSOL  
 CC OUT OF THE CELL OR INTO ORGANELLES (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + CA(2+)(CIS) = ADP + PHOSPHATE +  
 CC CA(2+)(TRANS)  
 CC -1- ENZYME REGULATION: ACTIVATED BY CALMODULIN (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- DOMAIN: THE N-TERMINUS CONTAINS AN AUTOINHIBITORY CALMODULIN-  
 CC BINDING DOMAIN, WHICH BINDS CALMODULIN IN A CALCIUM-DEPENDENT  
 CC FASHION (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (E1-E2 ATPASES). SUBFAMILY IIB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AB023045; BAB01709.1;  
 DR InterPro: IPR001757; E1-E2-ATPase.  
 DR InterPro: IPR001454; Hydrolyase.  
 DR InterPro: IPR000661; Na\_H\_K\_ATPase.  
 DR Pfam: PF00122; E1-E2-ATPase; 1.  
 DR Pfam: PF00702; Hydrolyase; 1.  
 DR PRINTS: PR00119; CATATPASE.  
 DR PROSITE: PS00154; ATPASE\_E1\_E2; 1.  
 KW Hydrolyase; Calcium transport; Transmembrane; Phosphorylation;  
 KW ATP-binding; Metal-binding; Magnesium; Calmodulin-binding;  
 KW Multigene family; Hypothetical protein.  
 FT DOMAIN 1 181  
 FT TRANSMEM 182 202  
 FT DOMAIN 203 220  
 FT TRANSMEM 221 241  
 FT TRANSMEM 242 369  
 FT TRANSMEM 370 389  
 FT TRANSMEM 390 426  
 FT TRANSMEM 427 444  
 FT DOMAIN 445 844  
 FT TRANSMEM 845 863  
 FT TRANSMEM 864 874  
 FT TRANSMEM 875 895  
 FT TRANSMEM 896 915  
 FT TRANSMEM 916 938  
 FT TRANSMEM 939 950  
 FT TRANSMEM 951 975  
 FT TRANSMEM 976 993  
 FT TRANSMEM 994 1015  
 FT TRANSMEM 1016 1025  
 FT TRANSMEM 1026 1047  
 FT TRANSMEM 1048 1073  
 FT DOMAIN 1048 1073  
 FT MOD\_RES 44 55  
 FT MOD\_RES 482 482  
 FT MOD\_RES 789 789  
 FT METAL 793 793

SO SEQUENCE 1073 AA; 11748 MW; 5AF6220EBD908069 CRC64;  
 Query Match 4.7%; Score 110; DB 1; Length 1073;  
 Best Local Similarity 22.8%; Pred. No. 5.9;  
 Matches 93; Conservative 51; Mismatches 118; Indels 146; Gaps 20;  
 QY 9 LFVLAITSAGHGRSD-----VNSLITKVTALKEAGSGNEALELGLK 55  
 DB 197 ILIIAAVTSIALGIKTEGLKGMIDGSIAPAVLVIVYVSDYRSIQ--FQNLDEK 254  
 QY 56 GEIOTKPDVGOATKI-----LGSVSAIGKNSGDPATRI----- 90  
 DB 255 RNIDLEVRGGRVTKISIDYVVDVYPIRLIGDQVPADGVLSIGSLAIDESSMTGESKI 314  
 QY 91 -----ISGC-----LDIVAGIATTEG-----GPGMGIGAVAS 118  
 DB 315 VHKDQSPFLMGSGKIVADVGNMLVTGVINTGMLVASISEDTGETPIQVRLNGLAT 374  
 QY 119 FVSSILFLTGSSKNSVAVIDALSKHDEALORHAGAKRPFASAFIQWKQOON 178  
 DB 375 FI-GIVGL-----SVALVIVLAL-----LVRYFTGTOTDNGATOFIKGTTSSID 418  
 QY 179 LTDS-----DLSTIANVP-----VYKFSNFICQ--LESRIIS--OGAATT 214  
 DB 419 IVDQVAFITFAVTVIVVAVVEGPIPLAVTLTLAISMKRMADKALVRLSACETMGSAATT 478  
 QY 215 SLSDAKRAVDFILLYCOLVYVRETLVLDLALYRKGAHVAASVANAENVKLEADTL 274  
 DB 479 ICSDKTGT-----LTLMQMTVV-ET-----YAGSKMDVA--DNPGLHPLTVA--- 519  
 QY 275 DFLHKLPE--QALIGAVVHP-----ISASETSKATILNTYTFYQ 312  
 DB 520 -----LISEGYAONTGTNIFHPKDGVEIISGPTERAKILSMAYKLGW 562  
 RESULT 4  
 HLTY\_ECOLI STANDARD; PRT; 1023 AA.  
 ID HLTY\_ECOLI  
 AC P09963;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE HEMOLYSIN, CHROMOSOMAL.  
 GN HLTA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J96 / SEROTYPE O4;  
 RX MEDLINE=85234404; PubMed=3891743;  
 RA Felmlee T., Pellett S., Welch R.A.;  
 RT "Nucleotide sequence of an Escherichia coli chromosomal hemolysin.";  
 RL J. Bacteriol. 163:94-105(1985).  
 RN [2]  
 RP SEQUENCE OF 1-44 FROM N.A.  
 RC STRAIN=2001;  
 RX MEDLINE=85258115; PubMed=3894051;  
 RA Nicoud J.-M., Mackman N., Gray L., Holland I.B.;  
 RT "Characterisation of HlyC and mechanism of activation and secretion  
 RT of haemolysin from E. coli 2001.";  
 RL FEBS Lett. 187:339-344(1985).  
 CC -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD  
 CC CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY  
 CC DEFINED.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING  
 CC CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC  
 CC ACTIVITY.  
 CC -1- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE  
 CC INVOLVED IN PORE FORMATION BY THE CYTOTOXIN.

```
CC -1- PTM: PALMITOYLATED BY HLVC. THE TOXIN ONLY BECOMES ACTIVE WHEN
CC MODIFIED.
CC -1- DISEASE: THE HEMOLYSIN OF E.COLI IS PRODUCED PREDOMINANTLY BY
CC STRAINS CAUSING EXTRAINTESTINAL INFECTIONS, SUCH AS THOSE OF THE
CC URINARY TRACT.
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL M10133; AAA23975.1; -.
DR EMBL X02768; CAA26546.1; -.
DR PIR A24433; LEECA.
DR InterPro: IPR001343; Hemlysn_Ca_bind.
DR InterPro: IPR003355; Rtx_N.
DR Pfam PF00353; hemolysinCbind_2.
DR Pfam PF0382; RTX; 1.
DR PRINTS; PR00313; CABDNNGRPT.
DR PROSITE; PS00330; HEMOLYSIN.CALCIUM; 4.
KW Hemolysins; Toxin; Cytolysis; Cytolextin; Repeat; Calcium;
KW Transmembrane; Lipoprotein; Palmitate.
FT TRANSMEM 237 259 POTENTIAL.
FT TRANSMEM 267 326 POTENTIAL.
FT TRANSMEM 364 410 POTENTIAL.
FT DOMAIN 723 869 16 x REPEATS, GLY-RICH.
FT REPEAT 723 728 1.
FT REPEAT 732 737 2.
FT REPEAT 741 746 3.
FT REPEAT 750 755 4.
FT REPEAT 768 773 5.
FT REPEAT 777 782 6.
FT REPEAT 786 791 7.
FT REPEAT 795 800 8.
FT REPEAT 806 812 9.
FT REPEAT 816 821 10.
FT REPEAT 825 830 12.
FT REPEAT 834 839 13.
FT REPEAT 843 848 14.
FT REPEAT 855 860 15.
FT REPEAT 864 869 16.
FT LIPID 563 563 PALMITATE (BY SIMILARITY).
FT LIPID 689 689 PALMITATE (BY SIMILARITY).
FT VARIANT 6 6 A->T (IN STRAIN 2001).
SO SEQUENCE 1023 AA; 109867 MW; 196D5C0A9A28B54D CRC64; .

Query Match 4.7% Score 108.5; DB 1; Length 1023;
Best Local Similarity 19.4%; Pred. No. 7.1;
Matches 82; Conservative 65; Mismatches 163; Indels 113; Gaps 17;

OY 24 SOVNSL--LRKVFALKFASGSENALEALGLGEIQTKP-DRVGOATKIIGSVGSA 78
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 199 NNWVSFOOLNKLGISVLNNT-----KHLNGVGNNKLQNPMLNDNIGAGLDTVSGILSA 250
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 79 LGK---LNSGDA---TKIISGCLDIYAGIATTFGGPGVGMGTAVASFPVSSILSLFTGSSA 132
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 231 ISASFILSNADDTGTAKAAG----VELTTKVLIANGVGISQYIIIAORAAAGLSSTASA 305
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 133 KNSVAVIDRALSK-----HRDEAIQRHAAGAKRDFAESSAFIOVMKOQSNLTPDS 182
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 306 AGLIASVTTLASPFLSFSLIADKFRANKIEIYSQRFKKLTGSDGSLLAPFKETGAIDA 365
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 183 DLSTIANVPYVKFSNFGLQLESRIKSOGAATTSLSDAKRAVDFILYCQLVYMRETLVD 242
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
DB 366 SLT-----RISTVLVASVSGIS-AAATTSLVGAPVSA-----LVGATVGIISG 407
      :||| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
OY 243 LAIILRKGNAEHVAASV-----ENNRNVKELAAITDLFIHLKILFDQALLIGAVHIPISAS 297
```

```

Db      408  IIEASKOAMFEHVASKMADVIAMERKCKCKNTFENGYDRHRAFLIE----- 453
Oy      298  ETSKALNTKTKFEGVDPDRPIGNRRYKFTNSWNTYSICSEAYGMWYFRGCSNVNRPN 357
Db      454  DNEFKILSYGNKREYSV-----ERSVLTIOQHMTL----- 482
Oy      358  IVSKMSGDFYTWMSDR--KLYITKHDOGWCKGTLDDEPDGQHMRFIDLRKGTWV 414
Db      483  --IGELAG---VTRNGDKTLSCGSYIDYVEEG---KRLKKPDEFOKOVFPLKGNIDLS 534
Oy      415  SSK 417
Db      535  DSK 537

RESULT 5
HLA_ECOLI
ID      HLA_ECOLI      STANDARD:      PRF: 1024 AA.
AC      P08715;
DT      01-JAN-1988 (Rel. 06, Created)
DT      01-JAN-1988 (Rel. 06, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      HEMOLYSIN, PLASMID.
GN      HLYA.
OS      Escherichia coli.
OG      Plasmid Inci2 PHLY152.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RN      SEQUENCE FROM N.A.
RA      Hess J., Wels W., Vogel M., Goebel W.;
RT      "Nucleotide sequence of a plasmid-encoded hemolysin determinant and
RT      its comparison with a corresponding chromosomal hemolysin sequence.";
RL      FEMS Microbiol. Lett. 34:1-11(1986).
RN      [2]
RN      PALMITOYLATION AT LYS-564 AND LYS-690.
RX      MEDLINE=95099325; PubMed=7801126;
RA      Stanley P., Packman L.C., Koronakis V., Hughes C.;
RT      "Fatty acylation of two internal lysine residues required for the
RT      toxic activity of Escherichia coli hemolysin.";
RL      Science 266:1992-1996(1994).
RN      [3]
RN      PALMITOYLATION AT LYS-564 AND LYS-690.
RX      MEDLINE=96404790; PubMed=8808931;
RA      Ludwig A., Garcia F., Bauer S., Jarchau T., Benz R., Hoppe J.,
RA      Goebel W.;
RT      "Analysis of the in vivo activation of hemolysin (HlyA) from
RT      Escherichia coli.";
RL      J. Bacteriol. 178:5422-5430(1996).
RN      -1- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
RN      CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
RN      DEFINED.
RN      -1- SUBCELLULAR LOCATION: SECRETED.
RN      -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
RN      CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
RN      ACTIVITY.
RN      -1- DOMAIN: THE THREE TRANSMEMBRANE DOMAINS ARE BELIEVED TO BE
RN      INVOLVED IN PORE FORMATION BY THE CYTOOXIN.
RN      -1- PTM: PALMITOYLATED BY HLYC. THE TOXIN ONLY BECOMES ACTIVE WHEN
RN      MODIFIED.
RN      -1- DISEASE: THE HEMOLYSIN OF E. COLI IS PRODUCED PREDOMINANTLY BY
RN      STRAINS CAUSING EXTRAINTestinal INFECTIONS, SUCH AS THOSE OF THE
RN      URINARY TRACT.
RN      -1- SIMILARITY: BELONGS TO THE RIX PROKARYOTIC TOXIN FAMILY.
RN      -----
RN      This SWISS-PROT entry is copyright. It is produced through a collaboration
RN      between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN      the European Bioinformatics Institute. There are no restrictions on its
RN      use by non-profit institutions as long as its content is in no way
RN      modified and this statement is not removed. Usage by and for commercial
RN      entities requires a license agreement (See http://www.isb-sib.ch/announce/).
```

	Query Match	4.7%	Score 108.5	DB 1	Length 1024;	
	Best Local Similarity	20.7%	Pred. No. 7.1,	Mismatches 69;	Conservative 52;	Gaps 12.
	Matches					
DQ	24 SDVSNLLKRVETALKEASGSGNEALEGLNGEIQTKP--DRVGGATRKILGSVGSA LSK	81	: : : :	:	:	:
DQ	200 NNVNNSFSQQMLTL-----GSVLSTNRKLHNGVENCKLQNPNLNIGAGLDFTVSGILS AISA	254	: : : :	:	:	:
QY	82 ---LMSGA---FKIISCGLDIVAGIATITFGGPVGICAVASFVSII SLTGTSSAKKS	135	: : : :	:	:	:
DQ	255 SFILSNMADPRTRTKAAAG----VELTTKVLIQNVKGKISQYLTIIRAQAQGLSTAAAAGL	309	: : : :	:	:	:
QY	136 VAAVIDRALSK-----HRDIAIQIRNAGAARDPFRESASFIOVMQOQNLDDSDLS	185	: : : :	:	:	:
DQ	310 IASAVTTLIAISPSESLADIADKERFRANKIEYSORFVKLLGYDGDSLLAAFHKELGAI DASLT	369	: : : :	:	:	:
QY	186 ITAANVPYKFESNFIFGOLESRIQSOGAATTSLSDARKAVDFILLYOQLVYMRETLVDLA I	245	: : : :	:	:	:
DQ	370 TI-----SYVASVSSGIS-AAATTSLVGAHVSA-----LVGAVTGIISGLE	411	: : : :	:	:	:
QY	246 LYRKNAGEHVASA V-----ENANRYNKELAADTLDFLHLKIPEQALLIGAVYPHISASE TS	300	: : : :	:	:	:
DQ	412 ASKOMGEHVASHAMKDVA LAEMEKKKGNFYFENGCIARNAHFLE-----DNF	457	: : : :	:	:	:
QY	301 KAILNTYKYFEGVDPVPRPIGNRRYKFTNSYMYT	333	: : : :	:	:	:
DQ	458 KILSOYNKEYSV-----ERSVLITOOHWDT	482	: : : :	:	:	:
RESULT	6					
FAS3_RHOFA		STANDARD;	PRT:	312 AA.		
ID	FAS3_RHOFA					
AC	P46375;					
DT	01-NOV-1995 (Rel. 32, Created)					
DT	01-NOV-1995 (Rel. 32, Last sequence update)					
DT	01-FEB-1996 (Rel. 33, Last annotation update).					
DE	HYPOTHETICAL 33.6 KDA PROTEIN IN FASCINATION LOCUS PRECURSOR (ORP3).					

CC				
DR	EMBL; Z29635; CAA82743.1; -			
KW	Hypothetical protein; Plasmid; signal.			
FT	SIGNAL	1	28	POTENTIAL.
FT	CHAIN	29	312	HYPOTHEETICAL. 33 6 KDA PROTEIN IN
FT				FASCINATION LOCUS.
SO	SEQUENCE	312 AA;	33574 MW;	849PFI168ABD4404F CRC64;

	Query Match	4.5%; Score 105.5; DB 1; Length 312;
	Best Local Similarity	22.2%; Pred. No. 2.4;
	Matches 70; Conservative 48; Mismatches 93; Indels 105; Gaps 16;	
QY	54 LKGEIQTKPDVVGATKLIGSVGSALGK-----LMSGDATKTIISGLDIVAGIAT	103
Dd	25 MSGALETDP-----BAVVYLADIGAHLFEKAALADPNRVINVGIREQLMG---VAGGLAM	77
QY	104 TFGGPVYGKGIQAVASF-VSSILSLFTGGSSAKNSAAVAVIDRLSKNRDEAIORHAAGAKRD	162
Dd	78 CGMPRV---VHTVAALFVERPLEDIKLNFAOODVGAVL-----VSMGASYD	120
QY	163 FAESSAFQIVWKKQOSNLTDLSII-----AANVP-----VYKFS	197
Dd	121 LSE-PAF-----SHFPGDITVIDSMPKNTVHPGHPOEADLLLESPLCGDRVYL--	170
QY	198 NFIOQLSRIIS-----OGAATTSLSDAKRAVDETILLYCOLVMRETLVLDLAILY----	247
Dd	171 ----RLSQVVRYPHAAVKGTSTPTIKYIGTRVYLAANGPCIDAVYSATISMDDVTLLYKATI	226
QY	248 RKGNAEHVASVANENARVNKEELADLTDLFLKLLPEQALIGAVYHPISASETS--KAALN	305.
Dd	227 RPFATGICAAVQAQVNNPNNVLY-----EPYLAGTSAHQYSSSLVSHPRHLLS	274
QY	306 Y-----TKTEGVDP 314	
Dd	275 LGVRRMEDRRHYGTPD 290	
 RESULT 7		
ID	AER_ECOLI	STANDARD; PRT; 506 AA.
AC	P50466;	
DT	01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	AEROTAXIS RECEPTOR.	
GN	AER OR AIR OR B3072.	
OS	Escherichia coli.	

QY	26	VNSLITKVEITALKASQSGNE-----AALEALEGLKEIOTKRDYRGVAKTIILGVSVA	78
Db	239	MNOHAASVKQNSAFASAAADKISTIASNAAVOGGAMATTVIKTMD-TADSTORIGTITSL	357
QY	79	LGIINSGDATKIISGCDIYAGIATTFGGPYGMCIGAVASVSSILSLFTGSSAKNSVAA	138
Db	358	INDI-----AFQYINILALNAVAEAAAGEGCKFAVVAGEVRLAS--RSANAMANDIRK	409
QY	139	VIDBALSKHREAIORHAAGAKKRFAPESASFIOVMKQOOSULTDSDLSIIANVPVYFFSN	198
Db	410	LIDSADSVQSGSOQVHAAGFTME-----DIVAQRKNT-----	443
QY	199	FIGULESRKISGA--ATTSLDARKAVDFILLYCQ--LYVMRETLVLAILYKKGNAEH	254
Db	444	--QLIQISHSTLEQADGILSTLTRAVDLNLTKQNAELVEESAGVSAMV-----KH	493
QY	255	VASAVEANA 262	
Db	494	RASRLIEDA 501	

```

01 RESULT      8
02 RRPO_PVXCP STANDARD:          PRI:   1456 AA.
03 ID           RRPO_PVXCP
04 AC P22591.
05 DT 01-AUG-1991 (Rel. 19, Created)
06 DT 01-AUG-1991 (Rel. 19, Last sequence update)
07 DT 20-AUG-2001 (Rel. 40, Last annotation update)
08 DE RNA REPLICATION PROTEIN (165 KDA PROTEIN) (ORF 1) [CONTAINS: RNA-
09 DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE].
10 OS Potato virus X (strain CP) (PVX).
11 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.
12 OX NCBI_TaxID=12184;
13 RN [1]
14 RP SEQUENCE FROM N.A.
15 RX MEDLINE=90364772; PubMed=2392880;
16 RA Orman B.E., Celnik R.M., Mendi A.M., Torres H.N., Mentaberry A.N.;
17 RT "Complete cDNA sequence of a South American isolate of potato virus
18 X.";
19 RL Virus Res. 16:293-306(1990).
20 CC -I- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
21 POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.
22 CC -----
23 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
24 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
25 CC the European Bioinformatics Institute. There are no restrictions on its
26 CC use by non-profit institutions as long as its content is in no way
27 CC modified and this statement is not removed. Usage by and for commercial
28 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
29 CC or send an email to license@isb-sib.ch).
30 CC -----
31 DR EMBL: M31541: AAA47178.1; -.
32 DR EMBL: X55802: CAA39324.1; -.
33 DR PIR: S14005; S14005.
34 DR InterPro: IPR000606; Viral_helicesel.
35 DR Pfam: PF01443; Viral_helicesel.1.
36 KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;
37 KW Transferrase.
38 FT NP_BIND 735 742 POTENTIAL.
39 SO SEQUENCE 1456 AA; 165301 MW; 489BA57EA070BD2E CRC64;
40
41 Query Match 4.4%; Score 103; DB 1; Length 1456;
42 Best Local Similarity 18.7%; Pred. No. 27;
43 Matches 104; Conservative 70; Mismatches 187; Indels 194; Gaps 26;
44
45 QY 52 EGAKETIKTRDPKRGQARKIIASGVS-----ALGLKNSG-DATFKIIISG---CLDIYA 99
46 Db 714 EYLTKGIESTLEPKRVAAVCVIHAGSGSKSHAIQKALKREIGKSDITVTLPNELRLDWSK 773
47 QY 100 GIATP-----FGS-----PYGMGIGAVASPVSSI-LSLFQGSSSAK 133

```

Db 774 KVPNTPEYMFYKALIGTGSIYIFDDYSKLPFGY-IALVSESTIKILLILGDSRQ 832  
 QY 134 NSVAVIDRALSKHDEAIQ-----RHAAGA-----KDFPSSAFIYVMAQOQSLTSD 183  
 Db 833 SVYHETSDASIRHLGPATEVFAKYCRYLNAATHRNKDLAN--MLGVSERTGTTEIS 889  
 QY 184 LSI-----IAANVP-----VYKFSNFIQLESRI-----SQGAA 212  
 Db 890 MSSEPLEVPLVPSDEKRRRLMTGRNDTFTYACCGLTFRKQIVLDHNTQVCSANVM 949  
 QY 213 TTSLSDAKRAVDFT-----LLYCOLVYMBETLLVDLAIYRKGNAEHV 255  
 Db 950 YSALSRAVDRIHFINTSANSFAWEKLDSTPYLKTFLSVENHAKLKE-----YEPREAPI 1005  
 QY 256 AS-----AYENANRVKELADLTLD-----FLHKLIP 282  
 Db 1006 KEPEPOTMCVENESEVLEEKKELEKFDREIHSDAHGHSNCVOTEDTTIQLFESHQAK 1065  
 QY 283 EOLIGAVYHPIASSETS-----ILNVTYFGVDPVPRIGNRKRYET 327  
 Db 1066 DETLMTATIDARKLTSNOESNREFLSKRDIGDLFLNMQAMGLPKPEPI-----FS 1118  
 QY 328 NSYNTYSICSEAYWGMNFMFGCSNVRNPINRVSMSDGFYTMENSDRRLYTRKHQGM 387  
 Db 1119 QEVWEA---CAHEVQSKYLSKSKCNLNGTVROS-----PDFDENKIMVFLKSQ-- 1164  
 QY 388 GWGLDEDPGQGHMRFIP-LRHGKYV-----SSKRNPMWMEYSSASGYIR 435  
 Db 1165 -WVTKVEVLG-----LPKIKPGQTIATYQOTVLMFGTWARVMRQAFQPEVPI- 1215  
 QY 436 SWENNPQOCHWSIT 450  
 Db 1216 NCEETPEDMSAMUS 1230

RESULT 9  
 ID FLAA\_PSEAE STANDARD; PRT; 393 AA.  
 AC P21184;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE FLAGELLIN TYPE A.  
 GN FLAA.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 1-5.  
 RC STRAIN=PAK;  
 RX MEDLINE=91072275; PubMed=2123866;  
 RA Totten P.A., Iori S.;  
 RT "Characterization of the type A flagellin gene from Pseudomonas  
 RT aeruginosa PAK.";  
 RL J. Bacteriol. 172:7188-7199(1990).  
 CC -I- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
 CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.  
 CC -I- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M57501; AAA63458.1; -  
 DR PIR: A37853; A37853.  
 DR InterPro: IPR001029; Flagellin\_C.  
 DR InterPro: IPR001492; Flagellin\_N.  
 DR Pfam: PF00700; Flagellin\_C; 1.

DR Pfam: PF00669; Flagellin\_N; 1.  
 DR PRINTS: PR00207; FLAGELLIN.  
 DR PRODOM: PD000316; Flagellin\_C; 1.  
 FT INIT MET  
 SQ SEQUENCE 393 AA; 39909 MW; 828B551543F73C29 CRC64;

Query Match 4.48; Score 102; DB 1; Length 393;  
 Best Local Similarity 21.38; Pred. No. 5.7;  
 Matches 71; Conservative 39; Mismatches 92; Indels 132; Gaps 13;

QY 11 IVALTSKHKRSDVNSLTKVETALKKAGSN-EAALALEGLKGIOTRPPRGVQAT 69  
 Db 71 ISLAQT-AEGALQOSTNLIQMRDLISQANGSNSDSERTALNGEAKLOKELDRISWT 129  
 QY 70 ----KIL-GSVGSALGLNSGDATKII-----SG 93  
 Db 130 TFGGRKLIDGSFVYASQVQVGA-ANETISVIDEMSAESLNGTYFKADGGAVTAATASG 188  
 QY 94 CLDIIVAGIATTFGGPYGM----- 111  
 Db 189 TVDIAIGI--TGSAAVNVKVDKMGNETAEQAARKIAAVNDANVGIGAFSDGDTISYSK 246  
 QY 112 ----GIGAVASFVSIISLFTGSSA-----KNSVAVID-----RALSKH 147  
 Db 247 AGKDGSAITSVASGVYADIGSTVGTAAGVAPSAFAKTNPTVAKIDISTAKLSRR 306  
 QY 148 ---RDEAIORHAAGAKRDEAFESSAFIYVMAQOQN-----LTQSDLSITAAANVP 193  
 Db 307 AGDRTTAIKQIDASVPISVANQNFDTINLKNIGENVSAARGRIEDTFAAETANL-- 364  
 QY 194 YKFSNFIGLESRIISOGAATTSLSDAKRAVDFTL 227  
 Db 365 -----TKNOYLQOAGTAITLQAANDLPQSVL 389

RESULT 10  
 ID RPO\_PVXHB STANDARD; PRT; 1456 AA.  
 AC Q07630;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE RNA REPLICATION PROTEIN (165 KDA PROTEIN) (ORF 1) [CONTAINS: RNA-  
 DE DIRECTED RNA POLYMERASE (EC 2.7.7.48); PROBABLE HELICASE].  
 OS Potato virus X (strain HB) (PVX).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potexvirus.  
 OX NCBI\_TaxID=73488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94015010; PubMed=8409947;  
 RA Querol M., van der Vlugt R., Goldbach R., Salazar L.F.;  
 RT "Gen. sequence of potato virus X strain HB.";  
 RL J. Gen. Virol. 74:2251-2255(1993).  
 CC -I- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN  
 CC POSSIBLY FUNCTIONS AS A ATP-BINDING HELICASE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X72214; CA51012.1; -  
 DR InterPro: IPR000606; Viral\_helicase1.  
 DR Pfam: PF01443; Viral\_helicase1; 1.  
 KW ATP-binding; Helicase; RNA replication; RNA-directed RNA polymerase;  
 FT NP\_BIND 735 742 ATP (POTENTIAL).  
 SQ SEQUENCE 1456 AA; 165138 MW; 10730A1EADAFAFBA CRC64;

```

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.,
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: CARBAMOYL-PHOSPHATE + ASPARTATE =
CC ORTHOPHOSPHATE + N-CARBAMOYLASPARTATE.
CC -1- PATHWAY: SECOND STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -1- SUBUNIT: CONTAINS CATALYTIC AND REGULATORY CHAINS.
CC -1- SIMILARITY: BELONGS TO THE ATCASSES/OTCASSES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001099; AAB91125.1; -
DR TIGR; AF0106; --
DR InterPro; IPR002029; Carbmyltransf_asor.
DR Pfam; PF00185; OTCace; 1.
DR Pfam; PF02729; OTCace.N; 1.
DR PRINTS; PR00100; AOTCACE.
DR PRINTS; PR00101; ATCACE.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR Pyrimidine biosynthesis; transferase; Complete proteome.
KW Pyrimidine biosynthesis; transferase; Complete proteome.
SQ SEQUENCE 299 AA; 33633 MW; 70F2FC2EBACD720 CRC64;

Query Match 4.4%; Score 101.5; DB 1; Length 299;
Best Local Similarity 22.6%; Pred.No. 4.3;
Matches 75; Conservative 38; Mismatches 122; Indels 97; Gaps 16;

QY 23 RSDVNSILTKYETKALKESSGSENALEALEG-IGCEIOTKRDQRQAKIKIGSVASALGK 81
Db 13 REDIYVILKKAEEDVDVARG--EKKRLIEKILGNLFEP---STRNSETFAMKR 65
QY 82 LNSGDATKILISGCDIVAGIATTEGPGYGMGIGAVASVSSILSLFTSSAKNSVAAYID 141
Db 66 L-----GGDV--INMAGASSI-----AKGETLATPTI 91
QY 142 RALSKHRDEAIQRHAAGAKRDPASSAFIYQW-----QOSNLTDSDLSITANVPYRK 195
Db 92 RVSSYCAIYIRHPLEGAARFAAENSSVPYINAGDGAGCHPTQTLIDL-----YT 142
QY 196 FSNFIQOLESITISGCAATTSLSDAK--NAVDFIILYCOLVYMR-----ETILV 241
Db 143 IKKEGGRID-----GITLALMGDLKYSTHTISLILKALALFDMRIYILISPEALALPEDITE 197
QY 242 DLALILYRGMEHVAASAVE--NANRVNKEADLTDFLHKL-----IPEQALIGA----- 289
Db 198 DVSNEIRARARLEVIYSIEDLVYTRICKERPRPDEEFT-RKYSGSYRTAETILKSAKDSMI 256
QY 290 VYHPI-----SASSETSKAILNTKYEGVP 313
Db 257 VMHPLPRVDEIHPSYDSTKHARYFOQSGFYGP 288

RESULT 12
HTR1_HALN1 STANDARD; PRT; 535 AA.
AC P33741; O9HPF6;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SENSORY RHODOPSIN I TRANSDUCER (HTR-1) (METHYL-ACCEPTING PHOTOTAXIS
DE PROTEIN 1) (MP-I).
GN HTR1 OR HTR1 OR VNG1659G.
OS Halobacterium sp. (strain NRC-1), and
OS Halobacterium halobium.
```



```

OC Archaeae; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091, 2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shroga J.,
RA Swartzel S., Weir D., Hall J., Dahl T.A., Welcl R., Goo Y.A.,
RA Lettner B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
RA Alm M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarina S.;
RA "Genome sequence of Halobacterium species NRC-1.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10; 350-377 AND 457-476.
RC SPECIES=H. halobium; STRAIN=FLXSR;
RX MEDLINE=93101637; PubMed=1465418;
RA Yao V.J., Spudich J.L.;
RT "Primary structure of an archaeobacterial transducer, a
RT methyl-accepting protein associated with sensory rhodopsin I.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11915-11919(1992).
CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC SENSORY RHODOPSIN I (SR-I) TO THE FLAGELLAR MOTOR. RESPONDS TO
CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005075; AAC19913.1; -
DR EMBL; L05603; AAA72315.1; -
DR PIR; A47190; A47190.
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPR000658; DUF5.
DR InterPro; IPR003660; HAMF.
DR Pfam; PF00672; DUF5; 1.
DR Pfam; PF00015; MCPs1gnal; 1.
DR SMART; SM00304; HAMF; 2.
DR SMART; SM00283; MA; 1.
KW Transducer; Photoreceptor; Transmembrane; Methylation.
KW INT. MET
FT 0
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 28 POTENTIAL.
FT DOMAIN 29 38 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 39 54 POTENTIAL.
FT DOMAIN 55 535 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 265 265 METHYLATION.
FT MOD_RES 272 272 METHYLATION.
FT MOD_RES 279 279 METHYLATION.
FT MOD_RES 463 463 METHYLATION.
FT MOD_RES 472 472 METHYLATION.
FT MOD_RES 472 472 METHYLATION.
SQ SEQUENCE 535 AA; 56544 MW; B9945E4F66A9D091 CRC64;

```

```

Db 330 ----TDIEQRTMLA--LNASTEARAGNADGDFSVYADEVKDLAE--ETQDANETA 381
OY 138 AVIDRALSKHRD--EAIORHAGAKRDESSAFIQVKQOSNLTDLSIIANVPYK 195
Db 382 AVEKVTQATEDVTASIQO---TRTRVESGS-----ETVESFLRDIRTIADSLA--E 428
OY 196 FSNFIGOLESRISQGAN-----TSLSDAKRAVDFTLLYCOLVYMETLLVLDL 244
Db 429 VNSNIDETQRTTSEQAEVTQSTANSVERVAGLSDDTTAL-----ASDAESA 474
OY 245 ILRKGNAEHVAIVEN-ANRKNELADTDLFLKLPEDALIGAVYHPI 294
Db 475 VIGQESAEELIASLEPQONTAVEDLOSRVASFTYATEDSETAGSVDEQV 525

RESULT 13
YABE_BACSU
ID YABE_BACSU STANDARD; PRT: 437 AA.
AC P37546;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 47.7 KDA PROTEIN IN MENS-KSGA INTERGENIC REGION.
GN YABE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D26185; BAA05275.1; -
DR EMBL; Z99104; CAB1816.1; -
DR Subtilist; BG10103; yabe.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 437 AA; 47710 MW; E1B9B7CDED7FF94A CRC64;

Query Match 4.3%; Score 100; DB 1; Length 437;
Best Local Similarity 19.7%; Pred. No. 8.9;
Matches 87; Conservative 70; Mismatches 164; Indels 120; Gaps 22;

```

DB 264 KGKREKVVQKGEKGLKHFEEVVKENGEVSRRELVEETAFQSKDKVAVGTQKS--SPKF 321  
 QY 292 HPIASASTSAIINRYKYFEVDPVPRICGNRRKFTNSYNTNYSIGSEALMGVMPRGCS 351  
 DB 322 ETVASGDSKSTVVS-----RSNESTGKVTWSTAVTAS--CSGCS 360  
 QY 352 -----NVA-NPNIRYSKMSDGFYTWENSDRRKLYTTKHQDQGMGQ-TLDEPDQDGHM 402  
 DB 361 GHTATGVTNKKNNPNNAKVIANDPVNPIIGSK-----VAVEGYAIADTQSAIKGN 412  
 QY 403 --RPIPLRHGKYVSSKRPN 421  
 DB 413 IDVFEPKSSAY-----RMGN 428

RESULT 14  
 Y808\_CHLPH STANDARD; PRT; 444 AA;  
 ID Y808\_CHLPH  
 AC 092798;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHEICAL PROTEIN CPN0808/CP1063/CPJ0808.  
 GN CPN0808 OR CP1063 OR CPJ0808.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_Taxid=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CML029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).  
 RL [2]  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Ueberback T., Berry K., Bass S.,  
 RA Linher K., Welman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Wain M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39."; Nucleic Acids Res. 28:1397-1406(2000).  
 RL [3]  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CW029 from USA."; Nucleic Acids Res. 28:2311-2314(2000).  
 RL [1]  
 RN [1]  
 RP SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CP579/TC0868 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AE001662; RAD18946.1; -  
 CC DR EMBL; AE002263; AAF38836.1; -  
 CC DR EMBL; AP002548; BAA99016.1; -  
 CC DR TIGR; CP1063; -  
 CC KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 444 AA; 44477 MW; 4EA73AC9EF15EDCC CRC64;  
 Query Match 4.3%; Score 100; DB 1; Length 444;  
 Best Local Similarity 24.7%; Pred. No. 9.1;  
 Matches 54; Conservative 34; Mismatches 89; Indels 42; Gaps 8;

QY 28 SLTTKVEETLKEASGENEALALELGLKEIOTKPPRVGQATITLGSVSLGLKNSGDA 87  
 DB 143 ALLGLVMTLMANAG-----ESMKASFGSQONAIISQVSAIPAIGFAIRQANHQA 193  
 QY 88 T-----KIISGCDIVAGIATTTGCGPYGMI-----GAVASFVSSILFTGSSAKN 134  
 DB 194 SATEAQAKOSLISGIYNIW----FTVSYGAGIFSAKAGATSAIKSASFKEFGASAG 248  
 QY 135 SVAAYIDRLSKRDEAIORHAGAK-RDEASSAFIYWKQOOSNLTDSLSITIAANVPV 193  
 DB 249 GAAS--KALTSASSSVQQTMASTAKAATTAASSAGSAATKAAANLTD-DMAAASKMAS 304  
 QY 194 YKSFNFIGOL-----ESRISOGAATTSLSDAKRA 222  
 DB 305 DGASKRSGGLFGEVLKPKPMSEKVSQGMNVKTQGARVA 343

RESULT 15  
 PMAS\_ARATH STANDARD; PRT; 948 AA;  
 ID PMAS\_ARATH  
 AC 095373;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE ATPASE 5, PLASMA MEMBRANE-TYPE (EC 3.6.3.6) (PROTON PUMP 5).  
 GN AHAS OR AT2G24520 OR F25P17.18  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RX MEDLINE=20083487; PubMed=10617197;  
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,  
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,  
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,  
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,  
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,  
 RA Copenhagen G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,  
 RA Salzberg S.L., Fraser C.M., Venter J.C.;  
 RT "Sequence and analysis of chromosome II of Arabidopsis thaliana."; Nature 402:761-768(1999).  
 RL [2]  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Axelzen K.B.;  
 RL Unpublished observations (NOV-2000).  
 CC -1- FUNCTION: THE PLASMA MEMBRANE H+ ATPASE OF PLANTS AND FUNGI  
 CC GENERATES A PROTON GRADIENT THAT DRIVES THE ACTIVE TRANSPORT OF  
 CC NUTRIENTS BY H+-SYMPORT. THE RESULTING EXTERNAL ACIDIFICATION  
 CC AND/OR INTERNAL ALKALINIZATION MAY MEDIATE GROWTH RESPONSES (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (IN) = ADP + PHOSPHATE +  
 CC H(+) (OUT).  
 CC -1- SUBUNIT: BINDS TO 14-3-3 PROTEINS. THE BINDING IS INDUCED BY  
 CC PHOSPHORYLATION OF THR-947. BINDING TO 14-3-3 PROTEINS ACTIVATES  
 CC THE H+ ATPASE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (E1-E2 ATPASES). SUBFAMILY I1A.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A WRONG  
 CC N-TERMINUS PREDICTED FROM THE GENOMIC SEQUENCE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AC06954; AAD23893.1; ALT\_SEQ.

DR HSBP: P11719; 1EUL.

DR InterPro: IPR001757; E1-E2-ATPase.

DR InterPro: IPR000695; HATPase.

DR InterPro: IPR001454; Hydrolase.

DR InterPro: IPR000661; Na\_H\_K-ATPase.

DR Pfam: PF00122; E1-E2-ATPase; 1.

DR Pfam: PF00702; Hydrolase; 1.

DR PRINTS: PRO0119; CATAPASE.

DR PRINTS: PRO0120; HATPASE.

DR PROSITE: PS00154; ATPASE\_E1\_E2; 1.

KW Hydrolyase; Hydrogen ion transport; Transmembrane; Phosphorylation;  
KW Acetylation; ATP-binding; Metal-binding; Magnesium; Multigene family;  
KW Hypothetical protein.

FT INIT\_MET 0

FT DOMAIN 1

FT TRANSMEM 61

FT DOMAIN 81

FT TRANSMEM 93

FT DOMAIN 114

FT TRANSMEM 243

FT DOMAIN 264

FT TRANSMEM 273

FT DOMAIN 291

FT TRANSMEM 643

FT DOMAIN 665

FT TRANSMEM 670

FT DOMAIN 693

FT TRANSMEM 709

FT DOMAIN 730

FT TRANSMEM 751

FT DOMAIN 772

FT TRANSMEM 784

FT DOMAIN 805

FT TRANSMEM 812

FT DOMAIN 833

FT TRANSMEM 834

FT DOMAIN 848

FT MOD\_RES 1

FT MOD\_RES 328

FT MOD\_RES 947

FT METAL 587

FT METAL 591

FT SITE 946

FT SITE 948

SEQUENCE 948 AA; 104608 MM; CBFCAB8AE817BICE CRC64;

Query Match

Best Local Similarity 4.3%; Score 99; DB 1; Length 948;

Matches 101; Conservative 64; Mismatches 189; Indels 126; Gaps 23;

QY 12 VLAITSAGKGRSDVNSLTK-VETALKEAGSGSENALEALGKELQTKPRDVGQATK 70

DB 428 ILMLCNCKEYVRKRVGVIDKFAERGLRSIAVAROEVL-----KKDAPGPGWQ 477

QY 71 ILG-----SVGSALGKUNSGDARKIISGCLDIAGIATTTGGVPGMGIGAVASF 119

DB 478 LVGLPLFPDRHDSAEITRRALNLGVNVMKTTG--DOLA-IGKETGRRLGMGTNNPSS- 533

QY 120 VSSILSLFTGSSAKNSAAV-IDRALSKHDEAIQHRHAGAKRDEFAESSAFIOVMKQSN 178

DB 534 -----SALLGQYKDSLSGLAPDELIEK-----ADGFAGVPEPHKYEIVHRLQQRN 579

QY 179 ----LTDSPSLIANNVPYKFSNFIGLESRISSGCAATTSLSDAKR-AVDPIILYQQLV 233

DB 580 HICGMYGTDGVN-----DAPALKRADL-----GIAVVDATDAARGASDIIVLEPGLS 625

QY 234 VARETLVLVLAITYRGNNEHVAASAVENANRVKKELAADTLDLFLHKLIPQALIGAVYHP 293

DB 626 VIISAVLTSRAITQRKKN--YTTVAVSTIRIVFGFMFIALIWQDFSPPMVLIALIND 683

QY 294 ISASETSK-----ALIN-----YTKYFGVP 313

DB 684 GTIMTISKDRMKRSPQPDGKULRDISFTGVLGVALMTVFFVWMKDSDFSNFGV- 742

QY 314 DVPRPIGNRRYKFTNSYMTYSICSEAYVGNWYFRCSSNVNRPDNR-----VSKMSDGF 367

DB 743 ---RPLSORPEQWMAALYQVSIISQALLFVTRSRMSVAECPLLLGAFVIAQLVATP 799

QY 368 YTM-ENSDRRLKLYITKHQGMGW-GTLDDPDGQGMRRFP---LRHG-KYMWSSKRWP 421

DB 800 IAVYAWMSFARI-----EGAGWAGVY----WLSTFTYIPDLKFGILRYVLGKAWLN 851

Search completed: May 8, 2002, 12:16:27  
Job time: 195 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 12:11:47 ; Search time 19.46 Seconds  
(without alignments)  
1761.488 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322

Sequence: 1 MILKHLPLFLVLAITSAKH.....SGYIRSNENPPGQGHMSIT 450

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2317	99.8	450	2 JC7371	toxin - jellyfish
2	992.5	42.7	463	2 JC7372	toxin-A - jellyfish
3	137	5.9	1222	2 G59100	hypothetical prote
4	117	5.0	833	2 S54776	transcriptional fact
5	116.5	5.0	423	1 DEBYMC	malate dehydrogena
6	116	5.0	530	2 C72291	methy1-accepting c
7	112.5	4.8	510	2 C84718	probable kinesin 1
8	112.5	4.8	964	2 E71460	probable outer mem
9	111.5	4.8	656	2 A72428	methy1-accepting c
10	111.5	4.8	1336	2 T18288	ABC transport prot
11	110.5	4.8	656	2 E72379	methy1-accepting c
12	110.5	4.8	661	2 G72316	methy1-accepting c
13	110.5	4.8	1024	2 S10056	hemolysin A - Esch
14	110	4.7	810	2 B84185	cytochrome-like pr
15	109	4.7	541	2 B83032	probable chemotaxi
16	109	4.7	612	2 B75282	probable acyl-CoA
17	108.5	4.7	659	2 G82365	methy1-accepting c
18	108.5	4.7	1023	1 LEECA	hemolysin A - Esch
19	106.5	4.6	1098	2 B70232	hypothetical prote
20	106	4.6	1975	2 B81192	hemagglutinin/hemo
21	105.5	4.5	312	2 C55578	hypothetical prote
22	105	4.5	506	2 G65095	aerotaxis receptor
23	105	4.5	506	2 A85968	aerotaxis sensor r
24	103	4.4	1456	2 S14005	hypothetical prote
25	102.5	4.4	460	2 T19111	hypothetical prote
26	102	4.4	1456	2 A37853	flagellin, 40k - p
27	102	4.4	1456	2 JQ2294	hypothetical 165.1
28	101.5	4.4	299	2 B69263	aspartate carbamoy
29	101.5	4.4	428	2 D81255	histidinol dehydro

30	101	4.3	490	2 T00404	probable cytochrom
31	101	4.3	536	1 A47190	transducer protein
32	101	4.3	536	2 B84318	Hrt1 transducer li
33	101	4.3	1520	2 B82274	conserved hypotnet
34	100.5	4.3	307	1 DEPGC	3-hydroxyacyl-CoA
35	100.5	4.3	548	2 S15910	hemolysin-releasin
36	100.5	4.3	909	2 A75337	exonuclease ShcC -
37	100	4.3	437	2 S66069	conserved hypotnet
38	100	4.3	444	2 E72031	conserved hypotnet
39	100	4.3	444	2 F86591	C7579 hypothetical
40	99.5	4.3	1510	2 T13634	probable minor tai
41	99	4.3	484	2 E84254	hypothetical prote
42	99	4.3	931	2 F84637	probable plasma me
43	98.5	4.2	379	2 G82107	flagellin core pro
44	98.5	4.2	795	2 S26712	hypothetical prote
45	98.5	4.2	1517	2 T13329	hypothetical prote

#### ALIGNMENTS

RESULT	1	Score	DB 2	Length	450
JC7371	toxin - jellyfish (Carybdea rastoni)	99.8%	Pred. No. 8.9e-155;		
C:Species: Carybdea rastoni					
C:Date: 17-Nov-2000	#sequence: revision 17-Nov-2000	#text: change 17-Nov-2000			
C:Accession: JC7371	PC7094				
R:Nagai, H.; Takawa, K.; Nakao, M.; Ito, E.; Miyake, M.; Noda, M.; Nakajima, T.	Biochem. Biophys. Res. Commun. 275, 582-588, 2000				
A:Title: Novel proteotoxic toxins from the box jellyfish (sea wasp) Carybdea rasto					
A:Reference number: JC7371					
A:Accession: JC7371					
A:Molecule type: mRNA					
A:Residues: 1-450 <NAG>					
A:Cross-references: DDBJ:AB015878					
A:Accession: PC7094					
A:Molecule type: Protein					
A:Residues: 39-55;56-70;196-210;250-267;268-279;309-325;363-377;378-382 <NA>					
C:Comment: This protein, a member of bioactive protein, has hemolytic activity.					
C:Keywords: hemolysis; Inflammation; toxin					
Query Match	99.8%	Score 2317;	DB 2:	Length 450;	
Best Local Similarity	99.8%	Pred. No. 8.9e-155;			
Matches 449;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
QY	1	MILKHLPLFLVLAITSAKHGRSDVNSLTKVETALKEASGSNEALEGLKEIQT	60		
DB	1	MILKHLPLFLVLAITSAKHGRSDVNSLTKVETALKEASGSNEALEGLKEIQT	60		
QY	61	KPDVVGQATKILGVSAGLKLNSGDATKIISGLDIYAGIATTFGGVGMGIGAVASV	120		
DB	61	KPDVVGQATKILGVSAGLKLNSGDATKIISGLDIYAGIATTFGGVGMGIGAVASV	120		
QY	121	SSILSLFGSSAKNSVAVIDRALSKHDEAIORHAAGAKKDFESSAFIOVMKOOSNLT	180		
DB	121	SSILSLFGSSAKNSVAVIDRALSKHDEAIORHAAGAKKDFESSAFIOVMKOOSNLT	180		
QY	181	DSDSLIIANPVPYKFSNFIOLESRSISOGAATSLSDAKRAVDFITLLYCOLVVRFTLL	240		
DB	181	DSDSLIIANPVPYKFSNFIOLESRSISOGAATSLSDAKRAVDFITLLYCOLVVRFTLL	240		
QY	241	VDLAILYRKGAHVAHAVENANRVNKLADTLDFLHLKILPEOALIGAVYHPISASSETS	300		
DB	241	VDLAILYRKGAHVAHAVENANRVNKLADTLDFLHLKILPEOALIGAVYHPISASSETS	300		
QY	301	KAILNTRKFGVPPVPRIGNRKRTFTNSYNTYSISCEAYMGVYMPGCSNVRPNIRV	360		
DB	301	KAILNTRKFGVPPVPRIGNRKRTFTNSYNTYSISCEAYMGVYMPGCSNVRPNIRV	360		
QY	361	SKMSDGFYTMENSDRRKLYIRKHDOGWGWTLDDBDPOGQHMRFILPHGKYVYSSKRW	420		
DB	361	SKMSDGFYTMENSDRRKLYIRKHDOGWGWTLDDBDPOGQHMRFILPHGKYVYSSKRW	420		

R. Okunaka, R. T.; Cloud, K.; Hampton, O.; Hoffmaster, A. R.; Hill, K. K.; Kelm, P.; Koehler, J. Bacteriol. 181, 6509-6515, 1999

QY , 4 KHL-----PWLFIVLAITSAKHGKRSVDNSLTKVETALKEASG-SNEALEALEGLK--- 55

```

Db 235 KHLQKVAW-----DVVQSGSPESISLPSEPLSSEDELEGKSTEAATFCRSLKKNV 286
QY 56 ---GEIQTKPRVGOATKILGVSALGKNSDATTITISGCLDIYAGIATTTGGPYGMG 112
Db 287 FRSGELEPIPD-----SLDDESETISERLD-----STHHGAVGAG 321
QY 113 -----IGAVAFVSSILSLFTGSSAKNSVAVIDR----- 142
Db 322 ECSTKESHIRKKSKKKKKKVALATSSDSASVYTSKAKN---ALVDSSESGAVAREED 378
QY 143 -----ALSKHDEAIOI---HAACAKRDPFAESSAFIQMKQOQSNLTDSDLS 185
Db 379 VDHRAEAEAOACSTEKHR-EAMQRLPEPHEESNESASNSAARIISDRSDSDVD 437
QY 186 IIAANPVYKFSNFIGULSRISOGAATSLSDAKRAVFIILLYCOLVYVRETLVDAI 245
Db 438 LQSA---VAQLEFIPDIOER---AATV-----IRRYRDLG- 469
QY 246 LYRKGNAEHVA-----SAVENANRVNKEADTLDFLKLIPDQALLGAVVHPTSASET 299
Db 470 LFEKKAOCVAIRFGKFSKE-----NKQIEKNVODFL-----SLTGIES 509
QY 300 SKAIIATYKTFGVDPVPRPIGN--RRYKFTNSYNTYISGSEAYMGVNFRCGSNVRNPN 357
Db 510 ADKLITYTRY---PEEKLITITLTKRHAFF-----RLHIGKGIARWKLIV---Y 551
QY 358 IRVSKMSD-----GFYTMENSDRKLYTTKHQGM 387
Db 552 YRAKIFDVNNYKGRYNEEDFKLKAHSHLGNMW 586

RESULT 5
DEBYMC
malate dehydrogenase (EC 1.1.1.37), cytosolic - yeast (Saccharomyces cerevisiae)
N:Alternate names: Protein AOE23; protein 00537; protein YOL126c
C:Species: Saccharomyces cerevisiae
C>Date: 30-Jun-1993 #sequence_revision 19-Jul-1996 #text_change 16-Jun-2000
C:Accession: S63444; S12937; A34986; S05770; S66823; S71982
R:Casamayor, A.; Khalid, H.; Balcells, L.; Aldea, M.; Casas, C.; Herrero, E.; Arino, J.
Submitted to the EMBL Data Library, November 1995
A:Description: Sequencing of a 13.4 kbp fragment of the left arm of chromosome XV reveal
A:Reference number: S63440.
A:Accession: S12937
A:Molecule type: DNA
A:Residues: 1-423 <CNS>
A:Cross-references: EMBL:U41293; NID:g1209710; PIDN:AAC49466.1; PID:g1209715
R:Minard, K.I.; McAlister-Henn, L.
Mol. Cell. Biol. 11, 370-380, 1991
A:Title: Isolation, nucleotide sequence analysis, and disruption of the MDH2 gene from S.
A:Reference number: S12937; MUID:91094852
A:Accession: S12937
A:Molecule type: DNA
A:Residues: 47-423 <MIN1>
A:Cross-references: EMBL:M62808; NID:g171915; PIDN:AAA34766.1; PID:g171916
A:Accession: A34986
A:Molecule type: protein
A:Residues: 48-73 <MIN2>
R:Kopetzki, E.; Entian, K.D.; Lotspeich, F.; Mecke, D.
Biochim. Biophys. Acta 912, 398-403, 1987
A:Title: Purification procedure and N-terminal amino acid sequence of yeast malate dehyo
A:Reference number: S05770; MUID:87185517
A:Accession: S05770
A:Molecule type: protein
A:Residues: 48-73, P, 75-81 <KOP>
R:Arino, J.; Casamayor, A.; Gamo, F.J.; Gancedo, C.; Iafuente, M.J.; Aldea, M.; Casas, C.
Submitted to the Protein Sequence Database, July 1996
A:Reference number: S66814
A:Accession: S66823
A:Molecule type: DNA
A:Residues: 1-423 <ART>
A:Cross-references: EMBL:274868; NID:g1420008; PIDN:CAA99145.1; PID:g1420009; GSPDB:GN00

```

```

A:Experimental source: strain S288C
R:Casamayor, A.; Khalid, H.; Balcells, L.; Aldea, M.; Casas, C.; Herrero, E.; Arino
Yeast 12, 1013-1020, 1996
A:Title: Sequence analysis of a 13.4 kbp fragment from the left arm of chromosome X
ending frames.
A:Reference number: S71978; MUID:97051588
A:Accession: S71982
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-423 <CANP>
A:Cross-references: EMBL:U41293; NID:g1209710; PIDN:AAC49466.1; PID:g1209715
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995
C:Genetics:
A:Gene: SGD:MDH2; MIPS:YOL126c
A:Cross-references: SGD:S0005486; MIPS:YOL126c
A:Map position: 15L
C:Superfamily: L-lactate dehydrogenase
C:Keywords: cytosol; homodimer; NAD; oxidoreductase
F:48-423/Product: malate dehydrogenase #status experimental <MAT>
F:228,261/Active site: Asp, His #status predicted
F:231/Binding site: substrate (Arg) #status predicted

```

```

Query Match 5.0%; Score 116.5; DB 1; Length 423;
Best Local Similarity 20.2%; Pred. No. 1.1;
Matches 72; Conservative 65; Mismatches 125; Indels 95; Gaps 17;

```

```

QY 13 LAITSARKHGRSDVNSL-TVEETALKEASGS-----NEALEALEGKGEIOTRPD 63
Db 62 IAILGAAGIGQSSLSLKLKAOLOYOLKESNRSVTHILALVDVNGEALINGVATDLSHDT 121
QY 64 RVGOATKILGVSALGKNGSDATKIISG-----CLDIYAGIATTTGGPYGMG 112
Db 122 PIVSSH--SPAGCIENCLHNASTIVIPAGVPRKPGMTRODLFVFNAGIISQLDSI--- 176
QY 113 IGAVAFVSSILSLFTGSSAKNSVAVIDRALSKH---RDEAIOIRHAG-AKRDFAESSA 168
Db 177 --AECCLSKFVLVINPNVSLPVAVNSLNKHKHPSRNSGIERRMGVTKLDIVRAST 234
QY 169 FIOVMKQOQSNLTD-----SDLSITAN-----VYVYKFSNFIGOL----- 203
Db 235 FLKRNIESGLTPRVNSMPVPVIGHSGEITILFQSNSFLSLNBDQLKYLHRYOG 294
QY 204 -----ESRISOGAATTSLSDA--KRAVDFILLYCOLVYVRETLVDAIILYRKGNAEHVA 256
Db 295 GDEVYAKANKKGSATISMAHAGKCVQEV-----SLTL-----GNIEQIH 335
QY 257 SA---VENANRVNKEADTLDFLKLIPDQALLGAVVHPTSASERSKAIIANTYX 309
Db 336 GTYYVPLKDNANPNPIAPGAD-----QLLP--LVGDGADYFAIPLTITTKGV-STVDY 383

RESULT 6
C72291
methyl-accepting chemotaxis protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: C72291
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: C72291
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-530 <ARN>
A:Cross-references: GB:AF001771; GB:AE000512; NID:g4981678; PIDN:AAD36219.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1143
C:Superfamily: methyl-accepting chemotaxis protein

```



Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
 C.M.  
 Nature 399, 323-329, 1999  
 A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: A72428  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-656 <ARN>  
 A:Cross-references: GB:AE001690; GB:AE000512; NID:94980496; PIDN:AAD35117.1; PID:9498050  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0023  
 C:Superfamily: probable methyl-accepting chemotaxis transducer

Query Match  
 Best Local Similarity 4.8%; Score 111.5; DB 2; Length 656;  
 Matches 63; Conservative 44; Mismatches 107; Indels 95; Gaps 10;

QY 24 SDVNSILTKVETALKESGSENALEALGL--KGEIQTKPDVGOATKIIGSVGSALGK 81  
 Db 393 NMSSALTEVTSVGEVVAASQNIKITODLTERSEAVTKAARG--TERVEAVGVINK 450  
 QY 82 LNSG-----DATKIISGLDIVAGIA-----TFGGVGMGIGAV 116  
 Db 451 LKGSARORDYRLRELVDSAKTIGETIVDTISSIAEQTNLALNAIEAARGEGRGFAVY 510  
 QY 117 ASVSSILSTFTGSSAKNSVAVIDRALSKHDEAIOHNAAGKDFPSSAFIQVMKQO 176  
 Db 511 ADEIRKL-----AEESORATEDIAKMLSSIRATIEHVENSGKEMFGVDEIAVMGEE 562  
 QY 177 SNLTDSDLSIIANVPYKFSNFIOLE---SRISOGAATTSLSDAKRAVDITILYCOLV 233  
 Db 563 -----VTKRFREILGRIEINSMIENTAATAOEGAA----- 594  
 QY 234 VMRETLVDLAILYKRGNAEHVAASVENANVKNELADTDLFKLIPEDALLIGAVYHP 293  
 Db 595 -----AEEMASMDVNTKI-VEGVESLNRMSLEIDOT----- 627  
 QY 294 ISASETSKA 302  
 Db 628 ESARVSEA 636

RESULT 10  
 T18288  
 ABC transport protein - slime mold (Dictyostelium discoideum) (fragment)  
 C:Species: Dictyostelium discoideum  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T18288  
 R:Loomis, W.F.  
 submitted to the EMBL Data Library, August 1996  
 A:Reference number: Z18857  
 A:Accession: T18288  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1336 <LOO>  
 A:Cross-references: EMBL:066526; NID:91513297; PID:91513298; PIDN:AAB06789.1  
 C:Genetics:  
 A:Gene: abca

Query Match  
 Best Local Similarity 4.8%; Score 111.5; DB 2; Length 1336;  
 Matches 66; Conservative 49; Mismatches 112; Indels 91; Gaps 14;  
 QY 7 PWLFIV-----LAITSAKI-----GKRDVNSLTKVETALKESGSENALEALGLGK 56  
 Db 334 PWFVLPYVWGLSTKVRHIPPYFDDEVDVRAITKAHDA-----SNRAPL-IICGLSK 386  
 QY 57 EIQT--KPDVGOATKIL-----GSVGSALGKLNSGDATKI--ISGLDIVAGIATTFG 106

Db 387 SYTKLFRKKTVHAKVYLSLSEVKGITLIGLGSNGCGKSTTIGMLTGLEPTAGDALYGG 446  
 QY 107 GPVGMGIGAVASFEVSSI-----LSLEF-----GSSAKNSVAVIDRA-L 144  
 Db 447 HVSINIAAVRRITTSVPOHDILMAEMTAREHQLFSELGIPAOERESQIOKVLDDQVRL 506  
 QY 145 SKHREDAIORHAAGAR-----DFAESSAFIQVMKQSN-- 178  
 Db 507 SKISNNLISITSGCKMRRLSVAICIGDPKIFPMDEPTTGVDSKRHLIDLKYSIKNDK 566  
 QY 179 ---LSDSDI-----STIAANVPYKFSNFIOLESRSISOGAATTSLSDAKRAVDITL 227  
 Db 567 VILTSHDMHEVELADKIVIMNEGVMCNGNSLOLKSXYGQSVNIYAKSPESIPAVY 626  
 QY 228 LYCOLV-----MRETL 240  
 Db 627 EFVTLSTPCCKFMKOSAL 644

RESULT 11  
 E72379  
 methyl-accepting chemotaxis protein - Thermotoga maritima (strain MSB8)  
 C:Species: Thermotoga maritima  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C:Accession: E72379  
 R:Nelson, K.E.; Clayton, R.A.; Gyll, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.;  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardsc  
 C.M.  
 Nature 399, 323-329, 1999  
 A>Title: Evidence for lateral gene transfer between Archaea and Bacteria from geno  
 A:Reference number: A72200; MUID:99287316  
 A:Accession: E72379  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-656 <ARN>  
 A:Cross-references: GB:AE001721; GB:AE000512; NID:94980922; PIDN:AAD35514.1; PID:94  
 A:Experimental source: strain MSB8  
 C:Genetics:  
 A:Gene: TM0429  
 C:Superfamily: probable methyl-accepting chemotaxis transducer

Query Match  
 Best Local Similarity 4.8%; Score 110.5; DB 2; Length 656;  
 Matches 62; Conservative 45; Mismatches 107; Indels 95; Gaps 10;

QY 24 SDVNSILTKVETALKESGSENALEALGL--KGEIQTKPDVGOATKIIGSVGSALGK 81  
 Db 393 NMSSALTEVTSVGEVVAASQNIKITODLTERSEAVTKAARG--TERVEAVGVINK 450  
 QY 82 LNSG-----DATKIISGLDIVAGIA-----TFGGVGMGIGAV 116  
 Db 451 LKGSARORDYRLRELVDSAKTIGETIVDTISSIAEQTNLALNAIEAARGEGRGFAVY 510  
 QY 117 ASVSSILSTFTGSSAKNSVAVIDRALSKHDEAIOHNAAGKDFPSSAFIQVMKQO 176  
 Db 511 ADEIRKL-----AEESORATEDIAKMLSSIRATIEHVENSGKEMFGVDEIAVMGEE 562  
 QY 177 SNLTDSDLSIIANVPYKFSNFIOLE---SRISOGAATTSLSDAKRAVDITILYCOLV 233  
 Db 563 -----VTKRFREILGRIEINSMIENTAATAOEGAA----- 594  
 QY 234 VMRETLVDLAILYKRGNAEHVAASVENANVKNELADTDLFKLIPEDALLIGAVYHP 293  
 Db 595 -----AEEMASMDVNTKI-VEGVESLNRMSLEIDOT----- 627  
 QY 294 ISASETSKA 302  
 Db 628 TSAKVSOA 636

RESULT 12  
 G72316



```

RESULT 13
S10056
hemolysin A - Escherichia coli plasmid pHLy152
C:Species: Escherichia coli
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 04-Mar-2000
C:Accession: S10056
R:Hess, J.; Wels, W.; Vogel, M.; Goebel, W.
FMS Microbiol. Lett. 34, 1-11, 1986
A:Title: Nucleotide sequence of a plasmid-encoded hemolysin determinant and its comparison with the
A:Reference number: S07209
A:Accession: S10056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1024 <HES>
A:Cross-references: EMBL:M14107
C:Genetics:
A:genome: plasmid pHLy152
C:superfamily: hemolysin A; hemolysin A homolog
C:Keywords: lipoprotein
F:247-792/Domain: hemolysin A homolog <HLyA>
F:564,690/Binding site: palmitate (Lys) (covalent) #status predicted

```

RESULT 14  
B84185  
cytochrome-like protein [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84185  
R:Ng, W.V.; Kennedy, S.P.; Mahiras, G.G.; Bergslut, B.; Pan, M.; Shukla, H.D.; Le  
; Lelthausen, B.; Keller, K.; Cruz, R.; Ganson, M.J.; Hough, D.W.; Maddocks, D.G.  
Jung, R.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID: 20504483  
A:Accession: B84185  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-810 <S>  
A:Cross-References: GB:AE004437; NID:g10579888; PIDN:AG18846.1; GSPDB:GN00138  
C:genetics:  
A:Gene: fbr

[illegible]

D<sub>b</sub> 342 DHDAEEGEGVELQ<sup>ET</sup>ITAGGSCVSDALSQVDTNLVAGIKALTTDFEAVILETGFRLAL 401

O<sub>Y</sub> 240 LVDLAILYRKGNAEHVASVE 260  
| | : ||| | : : ||  
402 LGDAREQYHRCNAERAAASVE 422

D<sub>b</sub>

RESULT 15

E83032

Probidie chemotaxis transducer PA4915 [imported] - *Pseudomonas aeruginosa* (strain PA01)  
C:Species: *Pseudomonas aeruginosa*  
:Data: 15-Sep-2000 #Accession: AF001728.1

```
CjDate: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
CjAccession: E83032
```

R;Stover, C.K.; Phadman, S.; Yuan, Y.

..; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A, Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. Reference number: NC006504. URL: 2004.02.22

A:Reference number: A82950; MUID:20437337  
A:Accession: E83032

A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-541 <STO>  
A;Cross-references: GB:

A:cross-references: GB:AE004004; GB:AE004091; NID:g9951184; PIDN:AGC08300.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 A:Concotions:

C;Genetics:  
A;Gene: PA4915

A;Gene: PA4915

Query Match	4.78;	Score 109;	DB 2;	Length 541;
Best Local Similarity	32.64;	Prod No 51;		

Best Local Similarity	22.6%;	Pred. NO. 5.1;
Matches	64;	Conservative 45; Mismatches 94; Indels 80; Gaps 13;

2 ILKHLPMILFIVLAITS-----AKHGKRSQ-VNSLITKVEMLKEASSGNEA 46

Db  
262 LRLHLAETARKLTITSPENSAISNOVSHVATSGNDTDMVATAVHOMSAVDVADLVDA 337

                  | | | |         | : : : : : : : : : :  
                  : | : | |         | : : : : : : : : : :

                  INQDNLVELELNEASGNEA 40

47 ALEALEGAKGETOT-----KDRBYGQATKTTGCGVCSATGVIYCCGATWETTTGGGCTTTT 66

[illegible]

322 AAASAESANSAHTGTGLVHANLDAIQGLSVWGEAGAVIDTLR--NKTDEISTVLEVIQ 379

99 -----AGIATTFGGPVMGIGAVASEVSS-----ILSLETGSS 131

380 NIAQQTNLALNAAIEARAGEAGSGFAVVADEVKSLATINHKATETIREMIEALQAGAS 439

132 AKNSVAVIDRALSKHDEA--TOR-HAAG-----AKRDEASSAFIQVMKQSNL 179

DbpA

440 SAVS-----MQSREDAQVSQRAHEAGALGLIAAVEGIAQSNAOISTATEEOTA 492

180 TDSLSIIANVPVYKFSNFIGOLESRISOGAATTSI.SPAKPA 222

493 TASEVS---ONTDSTNAS--TCE---VAEEQVMEGMSGNTTAA CAC  
 TASEVS | | | | | | | | | |  
 100 TDDDDSLIMNVEFVINSNF IGDDESKRISQGAFTTSLSPAKRA 222

493 TASEVS---QNDLSNLS--IGE---VAEGAVKTSTSSVELA 526

```
Search completed: May 8, 2002, 12:13:54
Job time: 127 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 12:11:27 ; Search time 15.85 seconds  
(without alignments)  
638.895 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322

Sequence: 1 MILKHPMLFIVLATSARK.....SGYIRSWENNPQGHWSIT 450

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/prodata/2/1aa/6C.COMB.pep.\*  
6: /cgn2\_6/prodata/2/1aa/6D.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	4.7	3031	1	US-07-689-008-2
2	100.5	4.3	3079	4	US-09-413-814-80
3	97	4.2	2123	4	US-08-968-685A-10
4	95	4.1	434	2	US-08-989-925-1
5	95	4.1	1456	1	US-08-803-973-2
6	95	4.1	1456	1	US-08-803-972-2
7	94.5	4.1	1489	6	5183745-2
8	94.5	4.1	1794	6	5183745-6
9	93.5	4.0	580	2	US-08-591-079-2
10	93	4.0	1296	1	US-08-480-604A-28
11	93	4.0	1296	2	US-08-405-496A-28
12	93	4.0	1296	4	US-08-915-136-28
13	92.5	4.0	432	4	US-08-705-660-18
14	92.5	4.0	432	3	US-08-989-045-18
15	92.5	4.0	699	3	US-08-694-865-16
16	92.5	4.0	699	3	US-09-124-491-16
17	92	4.0	497	4	US-08-740-223A-14
18	92	4.0	593	2	US-08-591-079-8
19	92	4.0	593	4	US-08-591-079-10
20	91.5	3.9	491	4	US-08-740-223A-13
21	91.5	3.9	937	1	US-08-253-155A-31
22	91	3.9	585	2	US-08-426-125-2
23	91	3.9	585	2	US-08-455-355-2
24	91	3.9	1026	2	US-08-542-003-6
25	91	3.9	1026	2	US-08-322-760A-6
26	90	3.9	619	1	US-07-762-132A-2
27	90	3.9	619	1	US-08-301-722A-4

28	89.5	3.9	15281	2	US-08-471-119A-2	Sequence 2, Appl
29	89	3.8	401	2	US-08-591-079-4	Sequence 4, Appl
30	89	3.8	803	1	US-08-156-232-10	Sequence 10, Appl
31	89	3.8	803	1	US-08-304-626-10	Sequence 10, Appl
32	89	3.8	803	1	US-08-316-301A-12	Sequence 12, Appl
33	89	3.8	803	2	US-08-611-928-10	Sequence 10, Appl
34	89	3.8	803	2	US-09-173-891-10	Sequence 10, Appl
35	89	3.8	803	4	US-09-076-137-12	Sequence 12, Appl
36	89	3.8	803	5	PCT-US92-03624-12	Sequence 12, Appl
37	88.5	3.8	879	1	US-08-220-151-2	Sequence 2, Appl
38	88.5	3.8	879	1	US-08-220-151-3	Sequence 2, Appl
39	88.5	3.8	879	1	US-08-413-118-2	Sequence 2, Appl
40	88.5	3.8	879	1	US-08-413-118-3	Sequence 2, Appl
41	88.5	3.8	879	1	US-08-413-118-10	Sequence 2, Appl
42	88.5	3.8	879	3	US-08-473-446-2	Sequence 106, App
43	88.5	3.8	879	3	US-08-473-446-3	Sequence 2, Appl
44	88.5	3.8	879	3	US-08-473-446-106	Sequence 3, Appl
45	88.5	3.8	1259	3	US-09-187-049-13	Sequence 13, Appl

#### ALIGNMENTS

RESULT 1  
US-07-689-008-2  
Sequence 2, Application US/07689008  
Patent No. 5268274  
GENERAL INFORMATION:  
APPLICANT: Ben-Bassat, Arie  
APPLICANT: Calhoon, Roger D  
APPLICANT: Fear, Anna L  
APPLICANT: Gelfand, David H  
APPLICANT: Meade, James H  
APPLICANT: Tal, Ronny  
APPLICANT: Wong, Hing  
APPLICANT: Benziman, Moshe  
TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE  
NUMBER OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McCutchen, Doyle, Brown & Eversen  
STREET: Three Embarcadero Center  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/689,008  
FILING DATE: 19910422  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 337,194  
FILING DATE: 12-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 496,236  
FILING DATE: 23-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy, Lisabeth Felix  
REGISTRATION NUMBER: 31547  
TELEPHONE: (415) 393-2000  
TELEFAX: (415) 393-2286  
TELEX: 340817 MACPAG SFO  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3031 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-07-689-008-2

Query Match 4.7%; Score 110; DB 1; Length 3031;  
Best Local Similarity 21.9%; Pred. No. 0.23;  
Matches 79; Conservative 43; Mismatches 127; Indels 112; Gaps 14;

QY 24 SDVSLTKVETALKEAS---GSENALEAL---EGIKGEIOTKPDVGOATKILGSVG 76  
DB 1423 TDESTITGVYAMITEAESPLASGRVIALISGGGGLNNLVOLLAQRKKA-KIGQDLV 1481  
QY 77 SALGKLSGATKIIISCLDIVAGIATTFGPGVGMGIG-----AVASEVSSIL 124  
DB 1482 LAHG---DOLTSYRSSPLTV-----GTVPLEMEPDWMYNNHRSRVIVGELLCIL 1529  
QY 125 SLFTGSSAKSVAAVIDRALSKHDEAIQRHAGAKRDPFAESSAFIO---VAKQSNLT 180  
DB 1530 -----IVAVMYRALAKH---ALRR---RRELEBERORTMNRVYLSLSCALL 1570  
QY 181 DSDSLITIAANYPVYKFSNFIGQLSRIISOGAATSLSDAKRAVDIFILYQGLVYMRFTL 240  
DB 1571 ASSCMYLVAVPVAR-----AQQASTAMTTAATSATRA-----PROILL 1609  
QY 241 VDLAILYKGNAEHVASAVENANRV-----NKEIADDTLDFLKL 281  
DB 1610 QQARFWLQOQOYDVARQALQNAERIANPSPVLEVEGYOTAIQNREAADTLRHLOQVA 1669  
QY 282 PEQALLGAVYHPIS-----ASETSKAILNTYTFGVDPVPRPIGNRRY 324  
DB 1670 PGSAAGNLNDLSERAISQSDLSQIRSLAGSGONAVAGYQKLFHGKPPHSLAVEY 1729  
QY 325 K 325  
DB 1730 Q 1730

RESULT 2  
US-09-413-814-80  
; Sequence 80; Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bloecker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413, 814  
; EARLIER FILING DATE: 1998-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 80  
; LENGTH: 3079  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-80

Query Match 4.3%; Score 100.5; DB 4; Length 3079;  
Best Local Similarity 22.5%; Pred. No. 2.3; Indels 113; Gaps 18;  
Matches 83; Conservative 47; Mismatches 126;

QY 38 KEASGSENALEALEGKGEIOTKPDR-----VGQATKILGSVG 76  
DB 55 REAAGAEVYWRKALDGTATIDLPDRRAHDAGARCRGAHAITLPKPLTGALARLBERG 114  
QY 77 SALGKLSGATKII---SCCLDIVAGI-----ATTFG-----VPGMIGVAVAS 118  
DB 115 TTFESVLLSALTYLLHRASQSDLVGVGPSGRHDESAFAFGYOMLVPRVALRGAAS 174  
QY 119 FVSEIISLFTGSSAKSVAAVIDRALS--KRRDEAIQRHAGAKRDPFAESSAFIOVAKQO 176  
DB 175 FDA-----LVANRDAFDALAHGSAHLHLARNGEAGQRALFVNAFAQ 221  
QY 177 SN--LDDSDSLITIAANYPVYKFSNFIGQLSRIISOG--AATSLSDAKRAVD--ILYQ 231  
DB 222 STEPSLDARLSALA-----IGVDVRIAGCELEITTLADEQAAAEFDLALFAE 270  
QY 232 L---VYMR---ETLVDLAILYKGNAEH---VASVENANRNKELADTDLFLKL 281  
DB 271 LDAGIALRFEYDQOLFEPATIERM--ARHFVLLSEVAEHPGRPLSELRM--LSDAERALL 327  
QY 282 -----PEQALLGAVYHPISASETSKAILNTYTFGVDPVPRPIGNRRY 325  
DB 328 LDDMSGAAARQAASAPACVHAHFBAHAAROPDATALER-----GHQRF 374  
QY 326 FTN-SYWN 333  
DB 375 YAOIWTST 383

RESULT 3  
US-08-968-685A-10  
; Sequence 10; Application US/08968685A  
; Patent No. 6214981  
; GENERAL INFORMATION:  
; APPLICANT: TUCKER, KENNETH  
; APPLICANT: PLOSILA, LAURA  
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968, 685A  
; FILING DATE: No. 6214961ember 12, 1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Baldwin, Geraldine F.  
; REGISTRATION NUMBER: 31,232  
; REFERENCE/DOCKET NUMBER: 7969-060  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2123 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-968-685A-10

Query Match 4.2% Score 97; DB 4; Length 2123;  
 Best Local Similarity 21.9%; Pred. No. 2.8; Mismatches 127; Indels 100; Gaps 16;  
 Matches 75; Conservative 40;

QY 42 GSNRAALEALGKELQTPDR--VGOATKILGVSALG--KL--NSGDATKI---I 91  
 DB 70 GISEA-----DGGKGGANMAGDKSIAIGDIAQALGSGSIAIGDKIYHNSNNNANITAKA 124  
 QY 92 SGCIDIYAGIATFGGPGVGIGCA-----VASFVSLSLFTGSSAKNSVAVIDRA 143  
 DB 125 SGNESTAIIGDVIYASGASIAISDILYKRETVQOISELLPIIRGOKALINDIYQADTN 164  
 QY 144 LSHRDEAIORHAAGAKRDEAFESSAFIOVMKQOONLTDSDLSITANVPYKKSNEIGQL 203  
 DB 185 LQRYRTHAGHASTA-----VGAMSYAKGHFSNAFGTR 218  
 QY 204 ESIISGAATSLSDAKRAVDILLYCOLVMBRELLVDAILIYRKGNAEHV---ASAVE 260  
 DB 219 ATAREGTYSIAVGLTATAKASSIAV-----GSMAQIGFAATAVG 258  
 QY 261 NARVVKELADTLDLHLKLIPEQALIGAVYHPIASSETSKALINTKYFGVDPVPRPIG 320  
 DB 259 GSIOVNLNGT-ALFGSGVOLKQNDVNA-----NVRAY--APDDNQPID 301  
 QY 321 NRRYK--FTNSYWNYSICSEAYMGNYMFRGCSNVNPNIRY 360  
 DB 302 N-RYKATFKNGATDVESI-----GN--SNGDSITRKILINV 334

RESULT 4  
 US-08-989-925-1  
 ; Sequence 1, Application US/08989925  
 ; Patent No. 5989820

## GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
 APPLICANT: Lal, Preeti  
 APPLICANT: Shah, Purvi  
 APPLICANT: Corley, Neil C.  
 TITLE OF INVENTION: HUMAN ADIPOPHILIN-LIKE PROTEIN  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Inocyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/989,925  
 FILING DATE: Herewith  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:

## FILING DATE:

ATTORNEY/AGENT INFORMATION:  
 NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0440 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-855-0555  
 TELEFAX: 650-845-4166  
 TELEX:

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 434 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear

IMMEDIATE SOURCE:  
 LIBRARY: BRAITUT13  
 CLONE: 1620223  
 US-08-989-925-1

Query Match 4.1% Score 95; DB 2; Length 434;  
 Best Local Similarity 24.5%; Pred. No. 0.31;  
 Matches 62; Conservative 35; Mismatches 94; Indels 62; Gaps 10;

QY 17 SANHGR-----SDVNSLITFVETALKESGSNEALEALGKELQTPDRVGOAT 69  
 DB 62 AAEKGVRTLLAAVSGAQPILSKLEPQIASASYAHGLDKLENNPILOQPTKEVLADT 121  
 QY 70 KILGVSALGKUNSGDARKIISGCDIYAGIATTFGGPGVGMIGAVASFVSSILSLFT- 128  
 DB 122 KEL--VSSKV---SG-AQEMVSSAKDTV---ATQISEAVDARGAVOSGVDTKSVYTG 171  
 QY 129 -----GSSAKNSVAVIDRALSKHDEAIORHAAGAKRDEAFESSAFIOVMKQOONLTD 182  
 DB 172 GVQSVAGSRIGQVNLGCVTLGK--SEWADNHL------LTDA 209  
 QY 183 DLSITANVPYKFSNF-----IGLESRIISGAATTSL-----SDAKRAVDPI 226  
 DB 210 ELARIATSLDGFDAVSQOQROEQSYFVRLGSLSERLROHAIHSLGLKLRATKORAEAL 269  
 QY 227 LLYCOLVMBRETL 239  
 DB 270 LQLSQALSIMETV 282

RESULT 5  
 US-08-803-973-2  
 ; Sequence 2, Application US/08803973  
 ; Patent No. 5773701

## GENERAL INFORMATION:

APPLICANT: Braun, Carl J.  
 APPLICANT: Hemenway, Cynthia L.  
 APPLICANT: Turner, Nigun E.  
 TITLE OF INVENTION: Plants Resistant to Infection by PVX  
 NUMBER OF SEQUENCES: 12  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. BBAF  
 STREET: 700 Chesterfield Village Parkway  
 CITY: St. Louis  
 STATE: Missouri  
 COUNTRY: USA  
 ZIP: 63198

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/803,973  
 FILING DATE: 21-FEB-1997  
 CLASSIFICATION: 800

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/804,862  
 FILING DATE: 04-DEC-1991  
 APPLICATION NUMBER: US 07/771,912  
 FILING DATE: 04-OCT-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Hoerner Jr., Dennis R.  
 REGISTRATION NUMBER: 30,914  
 REFERENCE/DOCKET NUMBER: 38-21(10541)A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (314)537-6099  
 TELEFAX: (314)537-6047

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1456 amino acids  
 TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-803-973-2

Query Match 4.1%; Score 95; DB 1; Length 1456;  
Best Local Similarity 19.0%; Pred. No. 2.4;  
Matches 105; Conservative 67; Mismatches 188; Indels 194; Gaps 27;

QY 52 EELKEIQTKPRVQCAATILISVGS-----ALGKINSQ-DATKIISG---CLDIYA 99  
DB 714 EYLTKIESLTERKVATCVIHAGSGSKSHAIQKALREIGKSDITVVLPTNELRLDMSK 773  
QY 100 GIATP-----FGG-----PYGMGIGAVASPVSSI-LSLFTGSSAK 133  
DB 774 KVPNTEPYWFKTEKALIGTGSIVIFDDYSKLPBGY-IEALICYSKIKVLILGDSHQ 832  
QY 134 NSVAVIDRALSKHDEAIQ-----RHAAGA---KRPFASSAFIOVMKQOONLTDSD 183  
DB 833 SYVHETAEADASIRHLGPATEYFSKYCRYLNTAHRNKKDLAN--MLGVYSERTGVTEIS 889  
QY 184 LSI-----IAANVP-----VYKFSNFIOLESRI-----SOGAA 212  
DB 890 MSAEFLGPIPLVPSDEKRLKMTGRNDFTYAGCGGLTKRKVOVLVDHNTQVCSANM 949  
QY 213 TSLSDAKRAVDPI-----LLYCOLVVMRETLVLDLAILYRKGAHEV 255  
DB 950 YTALSRAIDRIHFVNTSANSASFWEKLDSTPYLKTFLSVRRQALRE---YEPAAEPI 1005  
QY 256 AS-----AVENANRYKELADTLD-----FLHKLIP 282  
DB 1006 REPEPQTHMCVENESVLEEYKEELLEKEPDREIHSESHGSHNCVOTEDTTIOLFSHQAK 1065  
QY 283 EQLLGAVYH---PISASETS-----KALNTYKYGVPDVRPIGNRRYKFT 327  
DB 1066 DETLLMATIDARLKISNOTNREFRLSKDIDGVLFNTOKAMGLP-----KEPIPS 1118  
QY 328 NSYWNYSICSEAYMGNYMFRGCSNVNRPNIYVSKMSDGYTMENSDRRLKLYITKHQGM 387  
DB 1119 QEWMA---CAHEVQSKYLSKCNLINCTVROS-----PDEDEKIMVFLKSQ-- 1164  
QY 388 GNGTLEDDEGDGQHRFIP-LRHGKTMV-----SKRMPWEMYMESSASGIYR 435  
DB 1165 -WYTKYEKLG-----LPKIRPGQITAAFYQOTVLMFGTMARYMWFROAFOPKKEVFI- 1215  
QY 436 SWENNPGQGHWSI 449  
DB - 1216 NCETTPEDMSANAL 1229

RESULT 6  
US-08-803-972-2  
Sequence 2, Application US/08803972  
Patent No. 5792937  
GENERAL INFORMATION:  
APPLICANT: Braun, Carl J.  
APPLICANT: Hemenway, Cynthia L.  
APPLICANT: Turner, Nilgun E.  
TITLE OF INVENTION: Plants Resistant to Infection by PVX  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dennis R. Hoerner, Jr., Monsanto Co. B44F  
STREET: 700 Chesterfield Village Parkway  
CITY: St. Louis  
STATE: Missouri  
COUNTRY: USA  
ZIP: 63198  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/803, 972  
FILING DATE: 21-FEB-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/804, 862  
FILING DATE: 04-DEC-1991  
APPLICATION NUMBER: US 07/771, 912  
FILING DATE: 04-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Hoerner Jr., Dennis R.  
REGISTRATION NUMBER: 30,914  
REFERENCE/DOCKET NUMBER: 38-21(10541)A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314)537-6099  
TELEFAX: (314)537-6047  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1456 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-803-972-2

Query Match 4.1%; Score 95; DB 1; Length 1456;  
Best Local Similarity 19.0%; Pred. No. 2.4;  
Matches 105; Conservative 67; Mismatches 188; Indels 194; Gaps 27;

QY 52 EELKEIQTKPRVQCAATILISVGS-----ALGKINSQ-DATKIISG---CLDIYA 99  
DB 714 EYLTKIESLTERKVATCVIHAGSGSKSHAIQKALREIGKSDITVVLPTNELRLDMSK 773  
QY 100 GIATP-----FGG-----PYGMGIGAVASPVSSI-LSLFTGSSAK 133  
DB 774 KVPNTEPYWFKTEKALIGTGSIVIFDDYSKLPBGY-IEALICYSKIKVLILGDSHQ 832  
QY 134 NSVAVIDRALSKHDEAIQ-----RHAAGA---KRPFASSAFIOVMKQOONLTDSD 183  
DB 833 SYVHETAEADASIRHLGPATEYFSKYCRYLNTAHRNKKDLAN--MLGVYSERTGVTEIS 889  
QY 184 LSI-----IAANVP-----VYKFSNFIOLESRI-----SOGAA 212  
DB 1006 REPEPQTHMCVENESVLEEYKEELLEKEPDREIHSESHGSHNCVOTEDTTIOLFSHQAK 1065  
QY 283 EQLLGAVYH---PISASETS-----KALNTYKYGVPDVRPIGNRRYKFT 327  
DB 1066 DETLLMATIDARLKISNOTNREFRLSKDIDGVLFNTOKAMGLP-----KEPIPS 1118  
QY 328 NSYWNYSICSEAYMGNYMFRGCSNVNRPNIYVSKMSDGYTMENSDRRLKLYITKHQGM 387  
DB 1119 QEWMA---CAHEVQSKYLSKCNLINCTVROS-----PDEDEKIMVFLKSQ-- 1164  
QY 388 GNGTLEDDEGDGQHRFIP-LRHGKTMV-----SKRMPWEMYMESSASGIYR 435  
DB 1165 -WYTKYEKLG-----LPKIRPGQITAAFYQOTVLMFGTMARYMWFROAFOPKKEVFI- 1215  
QY 436 SWENNPGQGHWSI 449  
DB 1216 NCETTPEDMSANAL 1229

RESULT 7  
US-08-803-972-2  
Sequence 2, Application US/08803972  
Patent No. 5183745-2  
GENERAL INFORMATION:  
APPLICANT: DANCHIN, ANTOINE;GLASER, PHILIPPE;KRIN, EVELYN;

BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES  
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR  
; BIOLOGICAL USES  
; NUMBER OF SEQUENCES: 13  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,541  
; FILING DATE: 25-OCT-1989  
; SEQ ID NO: 2  
; LENGTH: 1489  
5183745-2

Query Match 4.1%; Score 94.5; DB 6; Length 1489;  
Best Local Similarity 25.0%; Pred. No. 2.8;  
Matches 65; Conservative 49; Mismatches 115; Indels 31; Gaps 12;

QY 11 IYLAITSARKGRSDVNSLITKVEETALKESGSENEALEALGLKEIO-----TKPDV 65  
DB 587 IALALAAAR-----CVTSGLOVAGASAGAAAGALAAALSPME-IYGLVQOSHAYADQDLK 640  
QY 66 GOATKILGVSALGKLNAGDATKIISGLDIYAGIATTFGGPVGMGIGA--VASEVSSI 123  
DB 641 AQESSAYGEEDALLAOLYRDKT-AAEGAVAGSAVISTVGAASVIAAASVGAAPAVY 699  
QY 124 LSFTGSSAKNSVAVIDRALSKH--RDEAIQRHAAGAKRDEFAESSAFIQVMQOQSLTD 181  
DB 700 TSLTGT--ALNGILRGVQOPIETKLANDYARKIDELGGPAYPEKN--LQARHEOLANSND 755  
QY 182 SDSLITANVPYKFSNFIOLESRIISOG-----AATTSLSDAKRAVD-FLLYCO-LVYM 235  
DB 756 GLRKMILADLOAGWNASSVIGVOTTEISKALIELAATGNADNLKSVDFVDFVQGERVA 815  
QY 236 RETLLVDA-----ILYRKG 250  
DB 816 GQPVLDVAAGIDIASRKG 835

RESULT 8  
5183745-6  
; Patent No. 5183745  
; APPLICANT: DANCHIN, ANTOINE; GLASER, PHILIPPE; KRIN, EVELYN;  
; BARZU, OCTAVIEN; LADANT, DANIEL; ULLMAN, AGNES  
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR  
; BIOLOGICAL USES  
; NUMBER OF SEQUENCES: 13  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/426,541  
; FILING DATE: 25-OCT-1989  
; SEQ ID NO: 6  
; LENGTH: 1794  
5183745-6

Query Match 4.1%; Score 94.5; DB 6; Length 1794;  
Best Local Similarity 25.0%; Pred. No. 3.9;  
Matches 65; Conservative 49; Mismatches 115; Indels 31; Gaps 12;

QY 11 IYLAITSARKGRSDVNSLITKVEETALKESGSENEALEALGLKEIO-----TKPDV 65  
DB 586 IALALAAAR-----CVTSGLOVAGASAGAAAGALAAALSPME-IYGLVQOSHAYADQDLK 639  
QY 66 GOATKILGVSALGKLNAGDATKIISGLDIYAGIATTFGGPVGMGIGA--VASEVSSI 123  
DB 640 AQESSAYGEEDALLAOLYRDKT-AAEGAVAGSAVISTVGAASVIAAASVGAAPAVY 698  
QY 124 LSFTGSSAKNSVAVIDRALSKH--RDEAIQRHAAGAKRDEFAESSAFIQVMQOQSLTD 181  
DB 699 TSLTGT--ALNGILRGVQOPIETKLANDYARKIDELGGPAYPEKN--LQARHEOLANSND 754  
QY 182 SDSLITANVPYKFSNFIOLESRIISOG-----AATTSLSDAKRAVD-FLLYCO-LVYM 235  
DB 755 GLRKMILADLOAGWNASSVIGVOTTEISKALIELAATGNADNLKSVDFVDFVQGERVA 814

QY 236 RETLLVDA-----ILYRKG 250  
DB 815 GQPVLDVAAGIDIASRKG 834

RESULT 9  
US-08-591-079-2  
; Sequence 2, Application US/08591079  
; Patent No. 5972899  
; GENERAL INFORMATION:  
; APPLICANT: Zychlinsky, Arturo  
; TITLE OF INVENTION: Apoptosis Induced by Shigella ipab  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/591,079  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Livnat, Shmuel  
; REGISTRATION NUMBER: 33,949  
; REFERENCE/DOCKET NUMBER: 15661-20017, 00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-0764  
; TELEFAX: (202) 887-0764  
; TELEEX: 90-4030 MRSNORSWSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 580 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-591-079-2

Query Match 4.0%; Score 93.5; DB 2; Length 580;  
Best Local Similarity 22.1%; Pred. No. 0.73;  
Matches 68; Conservative 47; Mismatches 114; Indels 79; Gaps 14;

QY 39 EASGSSEAL--EALGLKEIOTPD-----RVGATKILGVSALGKLNAGDATK 89  
DB 265 EESLNDLALFQSLQESKTEMERKSDYAAEVRAEELNRYMGCVKILGAL----- 317  
QY 90 IISGLDIYAGIATTFGGPVGMGIGANA--SFVSSILSFTGSS-----ANNSVAVIDR 142  
DB 318 -----LITVSVAFAAFSGASIALAAGLALMYDAIYQAAATGNSFEOALNPIKAVIE 372  
QY 143 ALSKRDDEAIQRHAAGAKRDEFAESSAFIQVMQOQSLTDLSLITANVPYKFSNFIO 202  
DB 373 PLIKLSDATFTKMEGLGVDSKRAKMGISIL--GAIAGALVLYAIVLVAT-----VGK 424  
QY 203 LESRISOGAATTSLSDAKRAVD-FLLYCO-LVYMRETLVLDLALLYRKGNAEHAASA 262  
DB 425 -----QAALAEINIGK-----ITGKT-LTDLIPKFLKNFSSQDLDTITNA 464  
QY 263 -NRVNEKLAADLDLPHKLI-----PEQALIG-----AYHPISASETSK 301  
DB 465 VARLKKFTLAGADEVYSKQIISTHLNQAVLLGESVNSATQAGGSVASAVFO--NSASTML 522  
QY 302 ALINTYTK 309

Db 523 ADLTSKY 530

RESULT 10  
US-08-480-604A-28

Sequence 28, Application US/08480604A

Patent No. 5736139

GENERAL INFORMATION:

APPLICANT: KINK, JOHN A.

APPLICANT: THALLEY, BRUCE S.

APPLICANT: PADHAY, NISHA V.

APPLICANT: FIRCA, JOSEPH R.

APPLICANT: STAPFORD, DOUGLAS C.

TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN &amp; CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/480,604A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/422,711

FILING DATE: 14-APR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/405,496

FILING DATE: 16-MAR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPD-01763

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:

LENGTH: 1296 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-480-604A-28

Query Match 4.0%; Score 93; DB 1; Length 1296;

Best Local Similarity 20.9%; Pred. No. 3.2;

Matches 112; Conservative 61; Mismatches 161; Indels 202; Gaps 28;

9 LFVLAITSKRGK-----RSDVNSL---TKVT-----ALKRSGSNEA--LEALEG 53

550 MFHLYRAQEFHSGSRIALFNSVNEALINPSRYVTFSSDYKVKVNAEAMFLGNVEQ 609

QY 54 L-----KGEIQKPRDVGQATKILGVSALGKLN SGD-----ATKIISG---L 95

Db 610 LVYDFDETSEVPTDKINDITIIIPYIGPA---LNIGMLYKDEFGALIFSGAVILL 665

QY 96 DIVAGIATTEGPGVGMGIGAVASVSSLSLFTGSSAKNSVAAV--IDRALSKHRD---- 149

Db 666 EFIEPIL-----PV-LGTFALVSYIA-----NKVLTVOTIDNALSKNEKDE 708

QY 150 -----EAIORHAGAKRDF-AESSAFIQVMKQSNLSDSD 183

Db 709 VYKVIYVNMILAKVNTQIDLIKKKKEALENOAEATKALINQYNOYTEEKNINIFNID 768

QY 184 L-----SIIANPVYKFSNFIG---QLESRIQGAATSLSDAKRAVDLFLYCOLV 234

Db 769 LSKNESINAKMININFLNOCSSVYLSNMIYGC-----VKRIEDF----- 811

QY 235 MREPLVDLAILYKRGNAEHVASAVEN-ANRPNKELADTDLFLKHLPEQALIGAVHP 293

Db 812 --DASLKDALLKLYINDNGLIGOVDRDKDKVNTLSTDIPQLSKYVNOHLLS----- 864

QY 294 ISASETSKAILNVTYTFGVDPVPRPIGNRRYKFTNSYNTVYSICSEAYMGNTYFGCSNV 353

Db 865 -TFTEYIKNIIIN-----TSILNLYRESNHL----- 888

QY 354 RNPNIKSKSDGFYTMENSDDRKLXITKHGOGMGWGLDEDPGOGHRRFIPLRHGXV 413

Db 889 ----IDLK-----YASKINIG---SKVNFPIDKNOQIENLESSEKIE 925

QY 414 VSKRMPNW-FMYMESSAGYIR-----SWENPPGQGHMSIT 450

Db 926 VILKNALVNSMYENFSTFWIRIPYFNSISLNNEYTIINCENNSG---WKVS 977

RESULT 11

US-08-405-496A-28

Sequence 28, Application US/08405496A

Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.

TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN &amp; CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/405,496A

FILING DATE: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.



```

?      REGISTRATION NUMBER: 40,027
?
?      REFERENCE/DOCKET NUMBER: OPND-01308
?
?      TELECOMMUNICATION INFORMATION:
?
?      TELEPHONE: (415) 705-8410
?
?      TELEFAX: (415) 397-8338
?
?      INFORMATION FOR SEQ ID NO: 28:
?
?      SEQUENCE CHARACTERISTICS:
?
?      LENGTH: 1296 amino acids
?
?      TYPE: amino acid
?
?      TOPOLOGY: linear
?
?      MOLECULE TYPE: protein
?
US-08-405-496A-28

```

OY 294 ISASETSKAILNRYKGVDPVPPRIGNRRYKFTNSWNTYSICSEAYMGVMRGCSNV 353  
: : : : :  
Db 865 -TFEYIKININ-----TSIINLRYESNHL----- 888  
OY 354 RNPRIYKSKMGDFYTMENSRRKLYITKHOGWGWTLDPPDQGMRIPIRHKYK 413  
: : : : :  
Db 889 ----IDSR-----YASKINIG---SKVNDPIDKNOIOIFNESSKIE 925  
OY 414 VSKRWPMW-FMYMESSAGYIR-----SWENPQPOGMSIT 450  
: : : : :  
Db 926 VILKNAIYVNMENFSTFWRIRIPKYFNSTLNNETIINCMMNSG----WKVS 977

## RESULT 13

US-08-705-660-18  
; Sequence 18, Application US/08705660  
; Patent No. 585683  
; GENERAL INFORMATION:  
; APPLICANT: KEESEE, SUSAN  
; APPLICANT: OBAR, ROBERT  
; APPLICANT: WU, YING-YJE  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: DETECTION OF CERVICAL CANCER  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Testa, Hurwitz & Thibault  
; STREET: 125 High St.  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/705,660  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GREENHALGH, DUNCAN A  
; REGISTRATION NUMBER: 38,678  
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 432 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-705-660-18

Query Match 4.0%; Score 92.5; DB 2; Length 432;

Best Local Similarity 23.9%; Pred. No. 0.56; Indels 43; Gaps 13;  
Matches 64; Conservative 44; Mismatches 117; Indels 43; Gaps 13;

OY 35 TALKEASGSNEALELEGKGEIOTKPDVGOATKILGSVSGALKRLNSGDATKIISGC 94  
: : : : :  
Db 12 SSISGSSGLGGSSRTSCRLSGGLGAGSCRGS-----GGIGSTLGGSSVSSCYSPGSG- 66  
OY 95 LDIYAGIATTFGPGVGMGIGAVASFVSSILSLFTGSSAKNSVAVID--PALSRRDEA- 151  
: : : : :  
Db 67 -----GGYSSFGVGDGLAGGEKA-----TMQNLNRLASYLDKVRALEAEANTELE 113  
OY 152 -----TORHAAGAKRPFASSAFIOVMKQOOSNLTDS-----LSITAAVPPYKF-SNF 199  
: : : : :  
Db 114 VKIRDWYQRCAPGPARDYSOYRTIELOKILATVDNANILLOIDNARLAADDFRTKF 173

OY 200 IGOLESISOGAATTSLSDAKRAVD-FILLYCOLVVRRETLVDLAILYRKGNAEHVASA 258  
: : : : :  
Db 174 ETEQALRLSVEA---DINGLARVLDLTLARADEMQIENKKEELA--YLKKNHEEMANA 228  
OY 259 VEN--ANRVKEL-ADPTLDFLKLPIE 283  
: : : : :  
Db 229 LRGOVGGEINVEDMADAGVD-LSRLINE 255

## RESULT 14

US-08-989-045-18  
; Sequence 18, Application US/08989045  
; Patent No. 6027905  
; GENERAL INFORMATION:  
; APPLICANT: KEESEE, SUSAN  
; APPLICANT: OBAR, ROBERT  
; APPLICANT: WU, YING-YJE  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: DETECTION OF CERVICAL CANCER  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Testa, Hurwitz & Thibault  
; STREET: 125 High St.  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/989,045  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GREENHALGH, DUNCAN A  
; REGISTRATION NUMBER: 38,678  
; REFERENCE/DOCKET NUMBER: MTP-023 (8395/27)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7000  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 432 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-989-045-18

Query Match 4.0%; Score 92.5; DB 3; Length 432;

Best Local Similarity 23.9%; Pred. No. 0.56; Indels 43; Gaps 13;  
Matches 64; Conservative 44; Mismatches 117; Indels 43; Gaps 13;

OY 35 TALKEASGSNEALELEGKGEIOTKPDVGOATKILGSVSGALKRLNSGDATKIISGC 94  
: : : : :  
Db 12 SSISGSSGLGGSSRTSCRLSGGLGAGSCRGS-----GGIGSTLGGSSVSSCYSPGSG- 66  
OY 95 LDIYAGIATTFGPGVGMGIGAVASFVSSILSLFTGSSAKNSVAVID--PALSRRDEA- 151  
: : : : :  
Db 67 -----GGYSSFGVGDGLAGGEKA-----TMQNLNRLASYLDKVRALEAEANTELE 113  
OY 152 -----TORHAAGAKRPFASSAFIOVMKQOOSNLTDS-----LSITAAVPPYKF-SNF 199  
: : : : :  
Db 114 VKIRDWYQRCAPGPARDYSOYRTIELOKILATVDNANILLOIDNARLAADDFRTKF 173  
OY 200 IGOLESISOGAATTSLSDAKRAVD-FILLYCOLVVRRETLVDLAILYRKGNAEHVASA 258  
: : : : :  
Db 174 ETEQALRLSVEA---DINGLARVLDLTLARADEMQIENKKEELA--YLKKNHEEMANA 228

OY 259 VEN--ANRVNKL--ADTLDFLKLPIE 283  
DB 229 LRGVGGEINVMADAPGD--LSRLINE 255

RESULT 15  
US-08-694-865-16

Sequence 16, Application US/08694865  
Patent No. 5837268

GENERAL INFORMATION:

APPLICANT: POTTER, ANDREW A.

APPLICANT: MANN, JOHN G.

TITLE OF INVENTION: GHRH-LEUKOTOXIN CHIMERAS

NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:

ADDRESSEE: REED & ROBINS LLP

STREET: 285 HAMILTON AVENUE, SUITE 200

CITY: PALO ALTO

STATE: CA

COUNTRY: USA

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/694,865

FILING DATE: 09-AUG-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: MCCracken, THOMAS P.

REGISTRATION NUMBER: 38,548

REFERENCE/DOCKET NUMBER: 9001-0016.22

TELEPHONE: (415)327-3400

TELEFAX: (415)327-3231

INFORMATION FOR SEQ. ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 699 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-694-865-16

Query Match 4.0%; Score 92.5; DB 2; Length 699;

Best Local Similarity 18.6%; Pred. No. 1.3;

Matches 93; Conservative 61; Mismatches 200; Indels 147; Gaps 21;

OY 24 SPVNSLITV-----ETLAKRSGSNEALALBGLKGIQTKPDVGQATKILGSVGS 77  
DB 219 SGISILGSVLAMDLDELALQNNNOHALAKAGLELTNSLI-----ENIANSVKTIDEPGE 274  
OY 78 ALGKINS-----GDATKTIISGC-----LDIVAGIAF-----TFGG 107  
DB 275 QISQSGSKLQNKIGLCTLDKDKNIGLGLDKAGLDLVISGLSGATAALYLVADKNASTAK 334  
OY 108 PYGNG-----IGAVASFVSSIL--SLFTSSAKNSVAAYIDRALSK----- 146  
DB 335 KYGAGFELANQVYGNITKAVSSYIIAQRYAAGLSSTGPAAALIASTVSLAISPLAFAGIA 394  
OY 147 ---HDEDAIQRAAGAKDFEASSAFIOVMKOOSNLTDSDLSIIANVPVYKFSNFIGOL 203  
DB 395 DKFNNAKSLSEYAEFFKRLGYDGNMLAEYRGCTGTIDASVTAI-----NTALAAI 445  
OY 204 ESRISOGAATSLSDAKRAVPFILLYCOLV-----MRETLVDLALILYRKG 250  
DB 446 AGGVSAALADLTFEKVKH-----NLVTNSKKEKVTIONMFERADFAKEVPNYKAT 496  
OY 251 NAEHVASAV--ENANRVNKLADTLDFLKLPIEQALIGAVYHPISASETSKAILNTYTKY 309  
DB 497 KDEKIEETIIGNGERITSKYVD-----LIAKNGKITDDELISKVYDNT--- 540

OY 310 FGVPDPPIGNRRYKFTNSYWNVTYSICEAYMGVYMRGCSNVRNPNIYRSKMSDGEFT 369  
DB 541 -----ELKHKNVNTNSLDKLISSV-----AFTSSNDSRNVLVAPTSMID----- 581  
OY 370 MENSDDRRLYITTKHDQGWGCTLDEDPDQGHMRFIPLRHGKYVWSSKRPMPWFMFMSS 429  
DB 582 ---QSLSLQFARGSQHWSYGL---RPGSGQDMSYGLRPGG---SSQHW-----SYGLRP 628  
OY 430 ASGYIRSWEN--NPGPGCHWS 448  
DB 629 GSG--SQDMSYGLRPGGSOHWS 648

Search completed: May 8, 2002, 12:13:31  
Job time: 124 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2002, 12:11:07, Search time 24.97 seconds

(without alignments)  
1334.921 Million cell updates/sec

Title: US-09-647-522-5

Perfect score: 2322

Sequence: 1 MILKHPLEFVLAITSAKH.....SGYIRSMENNPQGHMSIT 450

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2322	100.0	450	20	AAV33651
2	112.5	4.8	964	21	AA13634
3	112.5	4.8	964	22	AA13634
4	112.5	4.8	977	21	AA13640
5	112.5	4.8	977	22	AA13640
6	110	4.7	1066	21	AA13632
7	110	4.7	1073	21	AA13632
8	110	4.7	1086	21	AA13630
9	106	4.6	1978	20	AA13630
10	105.5	4.5	1844	18	AA13630
11	105.5	4.5	1844	19	AA13630

12	105.5	4.5	2504	18	AA13630
13	105.5	4.5	2504	19	AA13630
14	105.5	4.5	2504	20	AA13630
15	105.5	4.5	2504	21	AA13630
16	104.5	4.5	2504	22	AA13630
17	103.5	4.5	2504	23	AA13630
18	103.5	4.5	2504	24	AA13630
19	100.5	4.3	847	22	AA13630
20	100.5	4.3	847	23	AA13630
21	100	4.3	392	18	AA13630
22	99	4.3	885	21	AA13630
23	99	4.3	931	21	AA13630
24	99	4.3	956	21	AA13630
25	97.5	4.2	1013	19	AA13630
26	97.5	4.2	1149	22	AA13630
27	97.5	4.2	1445	12	AA13630
28	97	4.2	361	18	AA13630
29	97	4.2	361	18	AA13630
30	97	4.2	2123	22	AA13630
31	95	4.1	434	21	AA13630
32	95	4.1	568	22	AA13630
33	95	4.1	1456	14	AA13630
34	94.5	4.1	1522	10	AA13630
35	94.5	4.1	1705	10	AA13630
36	94.5	4.1	1705	10	AA13630
37	94	4.0	751	21	AA13630
38	94	4.0	798	21	AA13630
39	94	4.0	835	21	AA13630
40	93.5	4.0	835	21	AA13630
41	93.5	4.0	444	21	AA13630
42	93.5	4.0	444	21	AA13630
43	93.5	4.0	510	21	AA13630
44	93.5	4.0	580	21	AA13630
45	93	4.0	499	21	AA13630

#### ALIGNMENTS

RESULT 1	AAV33651	standard; Protein; 450 AA.
ID	AAV33651	
AC	AAV33651	
DT	06-JAN-2000	(first entry)
DE	C. rastonii hemolytic protein.	
KW	Hemolytic protein; blood platelet agglutination; drug development; treatment; sting; jellyfish; pharmaceutical; pesticide.	
OS	Carybdea rastonii.	
PN	MO950294	
PD	07-OCT-1999.	
PF	30-MAR-1999;	99WO-JP01607.
PR	01-APR-1998;	98JP-0088569.
XX	(SUNR) SUNTORY LTD.	
XX	Nagai H, Nakajima T;	
XX	WPI; 1999-580740/49.	
XX	N-PSDB; AA23610.	
PT	Protein with hemolytic activity, useful for drugs treating jelly fish stings; pharmaceuticals with blood platelet agglutination activity.	
PT	Pesticides by use of the hemolytic activity, and study of the hemolytic mechanism	

Photobacterium lumen  
Toxin TcBa, encode  
Photobacterium tcBa  
Modified Photobacterium  
Chlamydia trachomatis  
Escherichia coli H  
Escherichia coli H  
Botulinum toxin hea  
B. pertussis adeny  
Chlamydia pneumoniae  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Recombinant botulin  
Bacteriophage Dp-1  
N-terminal deleted  
H. pylori ORF 14ce  
H. pylori cytoplas  
Moraxella catarrhalis  
Human adipothelin  
Escherichia coli H  
PVX replicase. Po  
Sequence of part o  
Adenyl cyclase fro  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Arabidopsis thaliana  
Shigella flexneri  
Autoantigen diagno

xx Claim 4; Page 26-27; 32pp; Japanese.

cc This invention describes a novel protein which has hemolytic activity,  
cc blood platelet agglutination activity and a molecular weight of about  
cc 50,000 Da (by SDS-PAGE). The protein, homologs and analogs of the  
cc invention can be used as a novel approach to developing drugs useful for  
cc the treatment of jelly fish stings, pharmaceuticals with blood platelet  
cc agglutination activity, pesticides by use of the hemolytic activity, and  
cc in the study of the hemolytic mechanism. This sequence represents the  
cc hemolytic protein described in the invention

xx Sequence 450 AA;

Query Match 100.0%; Score 2322; DB 20; Length 450;  
Best Local Similarity 100.0%; Pred. No. 1.2e-195;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MILKHPMLFVLTATTSKAKHGKRSVNSLTFKVEFALKEASGSNEALEGLKEGIOT 60  
DB 1 MILKHPMLFVLTATTSKAKHGKRSVNSLTFKVEFALKEASGSNEALEGLKEGIOT 60  
OY 61 KPDVGVQATKILGVSAGALGKINSADAKKIISGLDIVAGIATTFGPGVGMGICAVASPV 120  
DB 61 KPDVGVQATKILGVSAGALGKINSADAKKIISGLDIVAGIATTFGPGVGMGICAVASPV 120  
OY 121 SSTLSLFGSSAKNSVAVIDRALSKHDEAIOHRAAKRPFARSSAFIOVMKQOSULT 180  
DB 121 SSTLSLFGSSAKNSVAVIDRALSKHDEAIOHRAAKRPFARSSAFIOVMKQOSULT 180  
OY 181 DSLSIIAANVPYKFSNFICGLESRIISOGAATSLSDAKRAVDLILLYCOLVWRETL 240  
DB 181 DSLSIIAANVPYKFSNFICGLESRIISOGAATSLSDAKRAVDLILLYCOLVWRETL 240  
OY 241 VDLALIRKGNAAHVASAVENANRYNKLADTDLFLKHLPEQLIGAVHPISASETS 300  
DB 241 VDLALIRKGNAAHVASAVENANRYNKLADTDLFLKHLPEQLIGAVHPISASETS 300  
OY 301 KAILWTVEFGVDPVPRPIGNRRYKFTNSYNTVSGISCFAYGNYMFGSCNVRNPRTV 360  
DB 301 KAILWTVEFGVDPVPRPIGNRRYKFTNSYNTVSGISCFAYGNYMFGSCNVRNPRTV 360  
OY 361 SKMSDFYTMENSDRRKLYITKHDGWSGTLDEDPDQGHMRFPLRHGKYMVSSKRP 420  
DB 361 SKMSDFYTMENSDRRKLYITKHDGWSGTLDEDPDQGHMRFPLRHGKYMVSSKRP 420  
OY 421 NMFYTMSSASGYIRSWENPGPGHWSIT 450  
DB 421 NMFYTMSSASGYIRSWENPGPGHWSIT 450

RESULT 2

AAB13634 AAB13634 standard; Protein; 964 AA.

xx AAB13634;

xx 02-FEB-2001 (first entry)

xx C. trachomatis pmpe gene protein.

xx Chlamydial infection; sexually transmitted disease;

xx pelvic inflammatory disease; PID; tubal obstruction; infertility;

xx trachoma; blindness; acute respiratory tract infection;

xx atherosclerosis; coronary heart disease; antibacterial.

xx Chlamydia trachomatis.

xx WO200034483-A2.

xx 15-JUN-2000.

PF 08-DEC-1999; 99WO-US29012.

XX 08-DEC-1998; 98US-0208277.

PR 08-APR-1999; 99US-0288594.

PR 01-OCT-1999; 99US-0410568.

PR 22-OCT-1999; 99US-0426571.

PA (CORI-) CORIXA COR.

XX (CORI-) CORIXA COR.

PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;

XX WPI; 2000-431303/37.

DR 2000-431303/37.

XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection

PT comprises immunogenic portion of Chlamydia antigen, which comprises

XX amino acid sequence encoded by polynucleotide sequence

PS Claim 2; Pages 184-186; 256pp; English.

XX The present invention relates to new nucleic acid sequences and the

CC proteins encoded by the nucleic acid sequences. The encoded proteins

CC comprise an immunogenic portion of a Chlamydia antigen. The encoded

CC proteins are useful for the serodiagnosis and treatment of Chlamydia

CC infection. Chlamydiae are intracellular bacterial pathogens that are

CC responsible for a wide variety of human infections. C. trachomatis

CC infection is one of the most common sexually transmitted diseases and can

CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction

CC and infertility. Trachoma due to ocular infection with C. trachomatis is

CC the leading cause of preventable blindness worldwide. C. pneumonia is a

CC major cause of acute respiratory tract infections in humans and is also

CC thought to play a role in the pathogenesis of atherosclerosis and

CC coronary heart disease. The present sequence is a protein isolated in the

CC present invention.

XX Sequence 964 AA;

Query Match 4.8%; Score 112.5; DB 21; Length 964;  
Best Local Similarity 21.3%; Pred. No. 0.66;  
Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

OY 21 GKRSDV-----NSLTKVTALKKESGSNEALEGLKEGIQKPDVGVQATKILGVS 76  
DB 254 gnrqnlvlynnrcfknvetasas-----dggalvltltdvlynnrgrlffs-- 301  
OY 77 SALCKLNSGDATFKIISGCLDIVAGIATTFGPGVGMGICAVASFVSIISLFGSSAKNSV 136  
DB 302 -----dnltnyggai---yapvvtlvnpgpyfnnlanng 336  
OY 137 AAVIDRALSKHDEAIOHRA-----AGAKR-----DFAESSAFIOV 172  
DB 337 gelyldgtsnakiad-dthallfneulvntnangstsnpprnatlvassgeall 395  
OY 173 MKQOS-NLTDS-LSIIAANVPYKFSNFICGLESRIISOGAATSLSDAKRAVD----- 224  
DB 396 gagsqnllydplevsnagvsv-sfnkeadqtsvsvsgatvnsadfhgrllqtkcpap 454  
OY 225 -----FILL--YCOLVVMRETL--LVDL---ATL--YRKGNAAHVASAVENARVKE 268  
DB 455 ltlngficiedhaglvtnrtftqygvsjngavlsckngtgd-----sasasatlkh 510  
OY 269 LAADTLDFLHLKLPBQALIGAVHPISASETSKALINT---KYGVDPVPRPI-----G 320  
DB 511 lgl-----lssllksgaei-----plwveplnnsnyladtaafsisdvlslddyg 561  
OY 321 NRRKFTNSYNTVSGISCFAYGNYMFGSCNVRNPRTVSKMSDGFTEMTSDRRKLYI 380  
DB 562 nspyesldl-----thalsq-----pmlstseasdnqjseidfsjlnv 602  
OY 381 TKHD-OG---MGWG-TLDEDPDQGHMR-----FIPL-RHGKYAV 414  
DB 603 phygwqglwtwqwtqtdpepassatldcpqkanfhtllltwlpagypvapkhrppl 662

OY 415 SSKRPWFMFMESSASGYIRSMENNPQGHMSIT 450  
 Db 663 antlwgmlateslkn-----saeltspgphfvgit 694

## RESULT 3

AAG83202  
 ID AAG83202 standard; Protein; 964 AA.

XX AAG83202;

XX 05-SEP-2001 (first entry)

XX Protein encoded by Chlamydia trachomatis pmpe gene.

KW Chlamydia; vaccine; infection; fusion protein; antigen; heart disease;  
 KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;  
 KM acute respiratory tract infection; Capl; CT529; OMCB;

XX polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.

XX Chlamydia trachomatis.

XX WO200140474-A2.

XX 07-JUN-2001.

XX 04-DEC-2000; 2000WO-US32919.

XX 03-DEC-1999; 99US-0454684.

XX 19-APR-2000; 2000US-0556877.

XX 20-JUN-2000; 2000US-0598419.

XX (CORI-) CORIXA CORP.

XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;

XX WPI; 2001-374831/39.

XX Chlamydia polypeptides and fusion proteins useful for preventing pelvic  
 PT inflammatory disease, trachoma, acute respiratory tract infections,  
 PT atherosclerosis and heart disease -

XX Claim 2; Page 191-193; 295pp; English.

XX The present sequence is provided in a specification relating to  
 CC compounds and methods for the treatment and diagnosis of chlamydial  
 CC infection. The compounds provided include polypeptides and fusion  
 CC proteins comprising immunogenic portions of Chlamydia antigens  
 CC and DNA sequences encoding such polypeptides. They are useful for  
 CC vaccinating against chlamydial infection, which causes pelvic  
 CC inflammatory disease, trachoma, acute respiratory tract infections,  
 CC atherosclerosis and heart disease.

XX Sequence 964 AA;

Query Match 4.8%; Score 112.5; DB 22; Length 964;  
 Best Local Similarity 21.3%; Pred. No. 0.66;  
 Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

OY 21 GKSDV-----NSLLTKVETLAKASGSNEALALEBGLKEIQTQPRVQAKRIILGSVC 76  
 Db 254 gngnvtvfyngnfknvetassas-----dggalkvtirldvtgnrgyrlffs-- 301  
 OY 77 SALGKINSQDATRIISGCLDIVAGIATTEGPGVMGIGAVAFSSILSLFTSSAKNSV 136  
 Db 302 -----dnltknnygal---yaprvltvdngptyflnniankng 336  
 OY 137 AAVIDRALSKHRDEAIORNA-----AGAKR-----DPAESSAFIOV 172  
 Db 337 galyldgtstnsklisa-drlaiflneivntvntangstsaamprrnatlvasssgelll 395  
 OY 173 MKQGS-NLTFSD-LSITAAVVPYKFSNFTGLESRTISQGAATTSLSDAKRAVD----- 224

Db 396 gaggsgnllfydplevsnaevsv-sfnkeadqsgsvvfgatvnsaafhnrlqtkcpap 454  
 OY 225 -----FILL--YQOLVMBETL--LVLD---AII--YRKGAHEHVASAVENANRYNKE 268  
 Db 455 ltsngflctedhaqltvtntftgtgvslyngavlsacykngcd-----sasnaslkh 510  
 OY 269 LAADTDLFLHKLPEQALLCAVYHPISASFTSKAILNT-----KYECVPVPRPI-----G 320  
 Db 511 lgin---lssllksgeel-----pllwpeptnmsnvtadtaafslsdvsksllddyg 561  
 OY 321 NRRYKFTNSYWNITYSICESEAYMGNYMFRGCSNVPNPIRVYSKMSDGYTTEMSDRRLYI 380  
 Db 562 nspyeatdl---thalsq-----pmlsiseasdnqjseidfsqlnv 602  
 OY 381 TKHD-QG--WGKG-TLDEPDGDOGHR-----FIPL-RHGKRYV 414  
 Db 603 phygwgglwvgyvaktqdpepassatitdpqkanrfrtllwlpagypspkhrppl 662  
 OY 415 SSKRPWFMFMFMESSASGYIRSMENNPQGHMSIT 450  
 Db 663 antlwgmlateslkn-----saeltspgphfvgit 694

## RESULT 4

AAB13640  
 ID AAB13640 standard; Protein; 977 AA.

XX AAB13640;

XX 02-FEB-2001 (first entry)

XX C. trachomatis pmpe gene amino terminus minus signal sequence protein.

XX Chlamydial infection; sexually transmitted disease;

XX pelvic inflammatory disease; PID; tubal obstruction; infertility;

XX trachoma; blindness; acute respiratory tract infection;

XX atherosclerosis; coronary heart disease; antibacterial.

XX Chlamydia trachomatis.

XX WO200034483-A2.

XX 15-JUN-2000.

XX 08-DEC-1999; 99WO-US29012.

XX 08-DEC-1998; 98US-0208277.

XX 08-APR-1999; 99US-0208594.

XX 01-OCT-1999; 99US-0410568.

XX 22-OCT-1999; 99US-0426571.

XX (CORI-) CORIXA CORP.

XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;

XX WPI; 2000-431303/37.

XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection

XX comprises immunogenic portion of Chlamydia antigen, which comprises

XX amino acid sequence encoded by polynucleotide sequence -

XX Claim 2; Pages 210-212; 256pp; English.

XX The present invention relates to new nucleic acid sequences and the  
 CC proteins encoded by the nucleic acid sequences. The encoded proteins  
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded  
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia  
 CC infection. Chlamydiae are intracellular bacterial pathogens that are  
 CC responsible for a wide variety of human infections. C. trachomatis  
 CC infection is one of the most common sexually transmitted diseases and can  
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction  
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is

CC the leading cause of preventable blindness worldwide. C. pneumonia is a  
CC major cause of acute respiratory tract infections in humans and is also  
CC thought to play a role in the pathogenesis of atherosclerosis and  
CC coronary heart disease. The present sequence is a protein isolated in the  
CC present invention.

XX Sequence 977 AA:

Query Match 4.8%; Score 112.5; DB 21; Length 977;  
Best Local Similarity 21.3%; Pred. No. 0.68;

Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

```
QY 21 GKRSDV-----NSLLRVERALKEAGSNEALALEGLKEIQTKPDRVGOATKILGSYG 76
DB 267 gnrgnlvyfncrfknvetasseas-----dgalkvtlrlidvngnrglffs-- 314
QY 77 SALGRLNSGDATKIISGCLDIYAGIATTFGPGVMGIGAVASVSSILSLFTGSSAKNSV 136
DB 315 -----dnltknysgal---yapvvlvdngpctyflnlanhky 349
QY 137 AAVIDRALSKHRDEAIQRA-----AGAKR-----DFAESSAFIOY 172
DB 350 gailyldgtsnksisa-drnallfneivnvnangtscsnprrnatltvasssgeill 408
QY 173 MKQOS-NLTDSO-LSIIAANVPYKFSNFIQLESRIISOGAATTSLSDAKRAVD----- 224
DB 409 gagssgnllfydpievsnagvsv-sfnkeadqtsvsvfsgatvnsadflqrmldqtktpap 467
QY 225 -----FILL--YCOLVVMRETL---LYDL---AIL--YRKGAHEVAVASAVENANRVNKE 268
DB 468 llsngflcliedhaqltvnrtftqgvsalngavlsckngtgd-----saasnaatlkh 523
QY 269 LAADTLDFLHKLIPQALIGAVYHPISASETSKAILNTY---KYRGVDPVPRPI-----G 320
DB 524 lglh-----lssllksgael-----pllwepclnsmnyadtaatslscvklsllddyg 574
QY 321 NRRYKFTSNYWNYSICSEAYMGNYMFRGCSNVNRPNIHVSKMSDGFYTMENSDRRKLYI 380
DB 575 nspysedtl---thalsqg-----pmlsiseasdnqgsenldfsglnv 615
QY 381 TKHD-OG---MGWG-TLDEDDPGDQGHMR-----FIPL-RHGKYWV 414
DB 616 phywgqglwtwvaktqdppepassatltdpqrkanfrhtllltwlpagyvspkhrspil 675
QY 415 SSKRWPWFMYESSASGYIRSMENNPQGGHMSIT 450
DB 676 antlwgmllateslkn-----saeltspgphfwgit 707
```

RESULT 5  
AAG83208  
ID AAG83208 standard; Protein; 977 AA.

XX AAG83208;

XX 05-SEP-2001 (first entry)

XX Protein encoded by Chlamydia trachomatis pmpe gene.

XX Chlamydia; vaccine; infection; fusion protein; antigen;  
XX pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;  
XX acute respiratory tract infection; Cap1; CT929; OMBB;  
XX polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.

XX Chlamydia trachomatis.

XX WO20010474-A2.

XX 07-JUN-2001.  
XX 04-DEC-2000; 2000WO-US32919.

PR 03-DEC-1999; 99US-0454684.  
PR 19-APR-2000; 2000US-0556877.  
PR 20-JUN-2000; 2000US-0598419.

PA (CORI-) CORIXA CORP.

XX Probst P, Bhatia A, Skelky YAW, Fling SP, Scholler J;

XX WPI: 2001-374831/39.

PT Chlamydia polypeptides and fusion proteins useful for preventing pelvic  
PT inflammatory disease, trachoma, acute respiratory tract infections,  
PT atherosclerosis and heart disease -

PS Claim 2; Page 216-218; 295pp; English.

CC The present sequence is provided in a specification relating to  
CC compounds and methods for the treatment and diagnosis of chlamydial  
CC infection. The compounds provided include polypeptides and fusion  
CC proteins comprising immunogenic portions of Chlamydia antigens  
CC and DNA sequences encoding such polypeptides. They are useful for  
CC vaccinating against chlamydial infection, which causes pelvic  
CC inflammatory disease, trachoma, acute respiratory tract infections,  
CC atherosclerosis and heart disease.

XX Sequence 977 AA:

Query Match 4.8%; Score 112.5; DB 22; Length 977;  
Best Local Similarity 21.3%; Pred. No. 0.68;

Matches 110; Conservative 61; Mismatches 184; Indels 161; Gaps 28;

```
QY 21 GKRSDV-----NSLLRVERALKEAGSNEALALEGLKEIQTKPDRVGOATKILGSYG 76
DB 267 gnrgnlvyfncrfknvetasseas-----dgalkvtlrlidvngnrglffs-- 314
QY 77 SALGRLNSGDATKIISGCLDIYAGIATTFGPGVMGIGAVASVSSILSLFTGSSAKNSV 136
DB 315 -----dnltknysgal---yapvvlvdngpctyflnlanhky 349
QY 137 AAVIDRALSKHRDEAIQRA-----AGAKR-----DFAESSAFIOY 172
DB 350 gailyldgtsnksisa-drnallfneivnvnangtscsnprrnatltvasssgeill 408
QY 173 MKQOS-NLTDSO-LSIIAANVPYKFSNFIQLESRIISOGAATTSLSDAKRAVD----- 224
DB 409 gagssgnllfydpievsnagvsv-sfnkeadqtsvsvfsgatvnsadflqrmldqtktpap 467
QY 225 -----FILL--YCOLVVMRETL---LYDL---AIL--YRKGAHEVAVASAVENANRVNKE 268
DB 468 llsngflcliedhaqltvnrtftqgvsalngavlsckngtgd-----saasnaatlkh 523
QY 269 LAADTLDFLHKLIPQALIGAVYHPISASETSKAILNTY---KYRGVDPVPRPI-----G 320
DB 524 lglh-----lssllksgael-----pllwepclnsmnyadtaatslscvklsllddyg 574
QY 321 NRRYKFTSNYWNYSICSEAYMGNYMFRGCSNVNRPNIHVSKMSDGFYTMENSDRRKLYI 380
DB 575 nspysedtl---thalsqg-----pmlsiseasdnqgsenldfsglnv 615
QY 381 TKHD-OG---MGWG-TLDEDDPGDQGHMR-----FIPL-RHGKYWV 414
DB 616 phywgqglwtwvaktqdppepassatltdpqrkanfrhtllltwlpagyvspkhrspil 675
QY 415 SSKRWPWFMYESSASGYIRSMENNPQGGHMSIT 450
DB 676 antlwgmllateslkn-----saeltspgphfwgit 707
```

RESULT 6  
AAG49832  
ID AAG49832 standard; Protein; 1066 AA.

AC AAG49832;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63081.  
XX  
KW Protein identification: signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP103405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140355.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149366.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.



```
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153750.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161992.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.
PR 29-OCT-1999; 990S-0162142.
```

Query Match 4.7%; Score 110; DB 21; Length 1066;

Best Local Similarity 22.8%; Pred. No. 1.3;

Matches 93; Conservative 51; Mismatches 118; Indels 146; Gaps 20;

```
OY 9 LFTVLTSAKKHGKRS-----VNSLLTKVETALKEAGSGNSNALEALEGLK 55
DB 190 iiliaavtstlaigkleglkegvlldggsiafavliviivtlaavsdgysiq--fgnIndex 247
OY 56 GEIQTKPRDVRVGATKI-----LGSVGSALGKLNSGDATKI-----90
DB 248 rmlqqlvmgmgtrtkvkslydvvvvgdvlpilrlgqvpadgylisghaladesmtgeski 307
OY 91 -----ISGC-----LDIVAGIATTFG-----GPVKGICAVAS 118
DB 308 vnhkdqspflmsgckvadvgvnmllvlgvgintewgllmasisedtgeetpilyrllnglat 367
OY 119 FVSSISLFTGSSAKNSVAVIDRALSKHDEAIOHAGAKRDPFESSAFIOMKQOSN 178
DB 368 fl-givgl-----svalvvlval-----lvryfgttcdtngatqfllkgtsisd 411
```

```
OY 179 LTDS-----DLSITIANVP-----YKFSNFIGO--LESRS----OGAATT 214
DB 412 ivddcvkiftiatliavtliavapeglplavtlttlysmrkmmadkalyrlisactmgsattc 471
OY 215 SLSDAKRAVDLILYCOLVYVRETLVDLAILYRKGAEHVASAVENANRVNELADTL 274
DB 472 lcsdkgct-----ltlmgmtvv-et-----yaggskmvda---dnpsglhpkiva--- 512
OY 275 DFLHKLPE---QALIGAVHP-----ISASETSKAILNTKRYGCV 312
DB 513 -----llsegvagntctgnlthpkdggveisgsplekallswayklgm 555

RESULT 7
AA649831
ID AA649831 standard; Protein; 1073 AA.
AC
XX AA649831;
AC
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 63080.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
XX 05-MAR-1999; 990S-0123180.
XX 09-MAR-1999; 990S-0123548.
XX 23-MAR-1999; 990S-0125788.
XX 25-MAR-1999; 990S-0126264.
XX 29-MAR-1999; 990S-0126785.
XX 01-APR-1999; 990S-0127462.
XX 06-APR-1999; 990S-0128234.
XX 08-APR-1999; 990S-0128714.
XX 16-APR-1999; 990S-0129845.
XX 19-APR-1999; 990S-0130077.
XX 21-APR-1999; 990S-0130449.
XX 23-APR-1999; 990S-0130510.
XX 28-APR-1999; 990S-0130891.
XX 30-APR-1999; 990S-0131449.
XX 30-APR-1999; 990S-0132048.
XX 30-APR-1999; 990S-0132407.
XX 04-MAY-1999; 990S-0132484.
XX 05-MAY-1999; 990S-0132485.
XX 06-MAY-1999; 990S-0132486.
XX 06-MAY-1999; 990S-0132487.
XX 07-MAY-1999; 990S-0132863.
XX 11-MAY-1999; 990S-0134256.
XX 14-MAY-1999; 990S-0134218.
XX 14-MAY-1999; 990S-0134219.
XX 14-MAY-1999; 990S-0134221.
XX 14-MAY-1999; 990S-0134370.
XX 18-MAY-1999; 990S-0134768.
XX 19-MAY-1999; 990S-0134941.
XX 20-MAY-1999; 990S-0135124.
XX 21-MAY-1999; 990S-0135353.
XX 24-MAY-1999; 990S-0135629.
XX 25-MAY-1999; 990S-0136021.
XX 27-MAY-1999; 990S-0136392.
XX 28-MAY-1999; 990S-0136782.
XX 01-JUN-1999; 990S-0137222.
XX 03-JUN-1999; 990S-0137528.
XX 04-JUN-1999; 990S-0137502.
```

PR	09-NOV-1999;	9905-0147416
PR	06-AUG-1999;	9905-0147493
PR	09-AUG-1999;	9905-0147935
PR	10-AUG-1999;	9905-0148171
PR	11-AUG-1999;	9905-0148319
PR	12-AUG-1999;	9905-0148541
PR	13-AUG-1999;	9905-0148684
PR	16-AUG-1999;	9905-0149368
PR	17-AUG-1999;	9905-0149175
PR	18-AUG-1999;	9905-0149426
PR	20-AUG-1999;	9905-0149722
PR	20-AUG-1999;	9905-0149723
PR	20-AUG-1999;	9905-0149829
PR	23-AUG-1999;	9905-0149902
PR	23-AUG-1999;	9905-0149930
PR	25-AUG-1999;	9905-0150566
PR	26-AUG-1999;	9905-0150864
PR	27-AUG-1999;	9905-0151065
PR	27-AUG-1999;	9905-0151066
PR	30-AUG-1999;	9905-0151303
PR	31-AUG-1999;	9905-0151438
PR	01-SEP-1999;	9905-0151930
PR	07-SEP-1999;	9905-0153363
PR	10-SEP-1999;	9905-0153700
PR	13-SEP-1999;	9905-0153758
PR	15-SEP-1999;	9905-0154018
PR	16-SEP-1999;	9905-0154039
PR	20-SEP-1999;	9905-0154379
PR	22-SEP-1999;	9905-0155133
PR	23-SEP-1999;	9905-0155486
PR	28-SEP-1999;	9905-0156458
PR	29-SEP-1999;	9905-0156596
PR	04-OCT-1999;	9905-0157117
PR	05-OCT-1999;	9905-0157553
PR	06-OCT-1999;	9905-0157865
PR	07-OCT-1999;	9905-0158029
PR	08-OCT-1999;	9905-0158232
PR	12-OCT-1999;	9905-0158369
PR	13-OCT-1999;	9905-0159293
PR	13-OCT-1999;	9905-0159294
PR	14-OCT-1999;	9905-0159329
PR	14-OCT-1999;	9905-0159331
PR	14-OCT-1999;	9905-0159637
PR	14-OCT-1999;	9905-0159684
PR	18-OCT-1999;	9905-0159588
PR	21-OCT-1999;	9905-0160741
PR	21-OCT-1999;	9905-0160767
PR	21-OCT-1999;	9905-0160768
PR	21-OCT-1999;	9905-0160814
PR	21-OCT-1999;	9905-0160815
PR	22-OCT-1999;	9905-0160980
PR	22-OCT-1999;	9905-0160981
PR	22-OCT-1999;	9905-0160985
PR	25-OCT-1999;	9905-0161404
PR	25-OCT-1999;	9905-0161405
PR	25-OCT-1999;	9905-0161406
PR	26-OCT-1999;	9905-0161359
PR	26-OCT-1999;	9905-0161360
PR	26-OCT-1999;	9905-0161361
PR	28-OCT-1999;	9905-0161920
PR	28-OCT-1999;	9905-0161932
PR	29-OCT-1999;	9905-0161933
PR	29-OCT-1999;	9905-0162142

Score 110; DB 21; Length 1073;

[illegible]

PR	07-MAY-1999	9905-01332663
PR	11-MAY-1999	9905-01342263
PR	14-MAY-1999	9905-01342128
PR	14-MAY-1999	9905-01334221
PR	14-MAY-1999	9905-01342211
PR	18-MAY-1999	9905-01343700
PR	19-MAY-1999	9905-01347168
PR	20-MAY-1999	9905-01334124
PR	21-MAY-1999	9905-01355553
PR	24-MAY-1999	9905-01336621
PR	25-MAY-1999	9905-01365692
PR	27-MAY-1999	9905-01336798
PR	01-JUN-1999	9905-01337222
PR	03-JUN-1999	9905-01375258
PR	04-JUN-1999	9905-01375702
PR	07-JUN-1999	9905-01337724
PR	08-JUN-1999	9905-01380924
PR	10-JUN-1999	9905-01385840
PR	10-JUN-1999	9905-01388447
PR	14-JUN-1999	9905-01336112
PR	16-JUN-1999	9905-01339452
PR	16-JUN-1999	9905-01339453
PR	17-JUN-1999	9905-01339461
PR	18-JUN-1999	9905-01339454
PR	18-JUN-1999	9905-01339455
PR	18-JUN-1999	9905-01339456
PR	18-JUN-1999	9905-01339457
PR	18-JUN-1999	9905-01339458
PR	18-JUN-1999	9905-01339459
PR	18-JUN-1999	9905-01339460
PR	18-JUN-1999	9905-01339461
PR	18-JUN-1999	9905-01339462
PR	18-JUN-1999	9905-01339463
PR	18-JUN-1999	9905-01339750
PR	18-JUN-1999	9905-01339753
PR	21-JUN-1999	9905-01339617
PR	22-JUN-1999	9905-01339689
PR	23-JUN-1999	9905-01403553
PR	23-JUN-1999	9905-01403554
PR	24-JUN-1999	9905-01406355
PR	28-JUN-1999	9905-01408233
PR	29-JUN-1999	9905-01409291
PR	30-JUN-1999	9905-01412827
PR	01-JUL-1999	9905-01412842
PR	01-JUL-1999	9905-01412843
PR	02-JUL-1999	9905-01420055
PR	06-JUL-1999	9905-01423290
PR	08-JUL-1999	9905-01428203
PR	12-JUL-1999	9905-01428220
PR	13-JUL-1999	9905-01433542
PR	14-JUL-1999	9905-01433624
PR	15-JUL-1999	9905-01440005
PR	16-JUL-1999	9905-01440085
PR	16-JUL-1999	9905-01440806
PR	19-JUL-1999	9905-01443325
PR	19-JUL-1999	9905-01443326
PR	19-JUL-1999	9905-01443332
PR	19-JUL-1999	9905-01443333
PR	19-JUL-1999	9905-01443334
PR	19-JUL-1999	9905-01443335
PR	19-JUL-1999	9905-01443336
PR	20-JUL-1999	9905-01443337
PR	20-JUL-1999	9905-01443338
PR	20-JUL-1999	9905-01443339
PR	21-JUL-1999	9905-01443340
PR	21-JUL-1999	9905-01443341
PR	21-JUL-1999	9905-01443342
PR	21-JUL-1999	9905-01443343
PR	21-JUL-1999	9905-01443344
PR	21-JUL-1999	9905-01443345
PR	21-JUL-1999	9905-01443346
PR	21-JUL-1999	9905-01443347
PR	21-JUL-1999	9905-01443348
PR	21-JUL-1999	9905-01443349
PR	21-JUL-1999	9905-01443350
PR	21-JUL-1999	9905-01443351
PR	21-JUL-1999	9905-01443352
PR	21-JUL-1999	9905-01443353
PR	21-JUL-1999	9905-01443354
PR	21-JUL-1999	9905-01443355
PR	21-JUL-1999	9905-01443356
PR	21-JUL-1999	9905-01443357
PR	21-JUL-1999	9905-01443358
PR	21-JUL-1999	9905-01443359
PR	21-JUL-1999	9905-01443360
PR	21-JUL-1999	9905-01443361
PR	21-JUL-1999	9905-01443362
PR	21-JUL-1999	9905-01443363
PR	21-JUL-1999	9905-01443364
PR	21-JUL-1999	9905-01443365
PR	21-JUL-1999	9905-01443366
PR	21-JUL-1999	9905-01443367
PR	21-JUL-1999	9905-01443368
PR	21-JUL-1999	9905-01443369
PR	21-JUL-1999	9905-01443370
PR		

PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 26-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 28-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 29-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match	4.7%	Score 110;	DB 21;	Length 1086;
Best Local Similarity	22.8%;	Pred. No. 1.3;		
Matches 93;	Conservative 51;	Mismatch		

9 LEFVLAITSAKHGKRS-----VNSTLTKEV----- matches 118; Indels 146; Gaps

```

210      |||aaavtslalgikitegllkegwldggstafav|||v|vrfrnec 55

```

```

36 GEIQTkEDRVGQAFKI-----LGSVGSALGKINSDDAMV*
      || : | | : ||
368 ----*--db                268 ----*--db

```

91 ----- 50

[illegible]

**119** EVSSII<sup>SLIETGCGAIVAA</sup> | : : : : | :  
| : : : : | :

[illegible]

```

-----svaIvIval-----lvryftgtgdtngatgflkyktsisd 178
179 LTDS-----DISTAANM

```

432 ivddcvkiftlavtlivvavncg)-----VYKRSNFIGQ--LESRIS-----QGAATP 214

215 SLSDAKRAVDILLYCOLVMMRETLIVR  
11  
-peg;riaveltlaysmrkmadkalvrlisacetsatt 491

[illegible]

275 DELHKLRE---QALIGAVYHR-----TSASETCVAVV---  
 | | | | |  
 yagsgskmdva---dnpsglhprkiva--- 532

```

533 -----| : ||
        |isegvaqnttgnlfhpkdgeveisspsfekattlsvv :
              ||||| :
                : |
```

20230727 9 575

AAV27230 standard; Protein: 1979 n

AA27230; 2510 AA.

24-SEP-1999 (first entry)

Amino acid sequence of *N. meningitidis* protein

protein UKF114-1.  
neisseria meningitidis protein; pharmaceutical: vaccine  
bacterial infection; treatment

*Neisseria meningitidis*.

W09936544-A2.

22-JUL-1999.

4-JAN-1999; 99WO-IB00103.

3-OCT-1998; 98GB-0022143.  
4-JAN-1998; 98GB-0000760.  
1-SEP-1999

1986; 98GB-0019015.

CHANNON SPA.



```

Db 713 ltlmalegfhvnglqghaslllaal-----kdgalvtvdagammkeesllqmaa 764
QY 158 G-AKRD-----PAESSAFIOVMKQOONLTDSDLSIIAANVPYKFS---NFIGOLESR1 207
Db 765 nqvekltkltswtqidaqlqwlqmsalavspidl--agmmalkygidhnyaaaw----- 817
QY 208 SGAATTSLSL-----AKRAVD--FLLYCOLVVMRETLVDLAILYRKGAHEHVASAVEN 261
Db 818 -qaaaaalmadhangqkklidelfskalcnyyl---navdsaaagvrdnnglytllidn 873
QY 262 -----ANRVKELADDTLDFLHKLIPEQALIGAVYHPISASETSKAILNTKRYGVPD 314
Db 874 qvsadvltstriaaiaaglylvnralnrdegqla-----sdvstrqfftdwery----- 922
QY 315 VPRPIGNRRKFTNSYVMNYISICSEA--YMGNYMFRGCSNVRNPNIRV--SKMSDGFYTM 370
Db 923 -----nkry-----stwgavselvyypeny-----dptgrlqgkmmadallqs 961
QY 371 EN-----SDRRKLYITKHDGKGWGTL-----DEDPGOGHMRFTPL----- 407
Db 962 lngsqlnadltvedafktyltstfeqvanlkvlsayhnhvndqgltyfigidqaapgytw 1021
QY 408 -----RHGKYVWSSKRMPNMFMYESSASGYIRSMEN 439
Db 1022 rsvdshkscengkf--aanaagew-----nkltcavnpwkn 1054

```

## RESULT 11

AAW56558  
ID AAW56558 standard; Protein; 1844 AA.

AC AAW56558;

DT 07-AUG-1998 (first entry)

DE Toxin TcdAII, encoded by the tcdA gene from genomic region tcd.

XX Photorhabdus luminescens W-14; nematode; symbiotic;

KW Heterorhabdus; tca; tcb; tcc; tcd; insecticidal activity; toxin;

KW Lepidoptera; Coleoptera; Hymenoptera; Diptera; Dictyoptera; Acarina;

KW Homoptera; Southern; Western corn rootworm; Colorado potato beetle;

KW mealworm; boll weevil; turf grub; beetle armyworm; black cutworm;

KW cabbage looper; codling moth; corn earworm; European corn borer;

KW tobacco hornworm; budworm.

XX Photorhabdus luminescens.

XX W09080932-A1.

XX 05-MAR-1998.

XX 05-MAY-1997; 97MO-US07657.

XX 06-NOV-1996; 96WO-US18003.

XX 28-AUG-1996; 96US-0705484.

XX 06-NOV-1996; 96US-0743699.

XX (DOMC) DOWELANCO.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Blackburn MB, Bowen DJ, Cliche JA, Ensign JC, Fatig R;

XX French-Constant RH, Guo L, Hey TD, Merlo DJ, Orr GU;

XX Petrelli J, Roberts JL, Rocheleau TA, Schoonover S;

XX Strickland JA, Sukhplinda K;

XX WPI, 1998-179427/16.

XX N-PSDB; AAW29986.

XX Isolated toxins from Photorhabdus luminescens strains - useful for

XX control of insect pests

XX Claim 34; Pages 254-259; 321pp; English.

```

CC The present sequence represents a protein named TcdAII of the bacterium
CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
CC nematodes of the Heterorhabdus genus. The bacterium has at least 4
CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
CC produced from these regions that are associated with insecticidal
CC activity. Peptides AAW56560 and AAW56562-65 are fragments of the present
CC protein. The native toxins are secreted proteins. The proteins are
CC toxic to insects upon exposure and especially when ingested. The
CC nucleic acid sequence can be used to produce transgenic plants,
CC baculoviruses or microbial hosts for toxin production. They can be used
CC to control insects pests from the Lepidoptera, Coleoptera, Hymenoptera,
CC Diptera, Dictyoptera, Acarina or Homoptera orders, especially the
CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,
CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,
CC codling moth, corn earworm, European corn borer or tobacco hornworm
CC or budworm.
CC
SQ Sequence 1844 AA:

```

Query Match 4.5%; Score 105.5; DB 19; Length 1844;  
Best Local Similarity 19.2%; Pred No. 7.2; Mismatches 184; Indels 151; Gaps 27;  
Matches 100; Conservative 85;

```

QY 11 IYLAITSKHKGRS---DVNSLTKRVETALKKASGSNEALEAL-----EG 53
Db 595 llatstslhgkesllgdelkramapcftsalhtc-segevaydlldlwdlqpaqltvog 653
QY 54 LKGEIOTKPRDR-----GQATKILGVSALGKLSGDKATKIIISCLDIVAGIATTFGP 108
Db 654 fweevqtptslkvltftqvlqaglslyrlg-lsetelslvtqssllvagksllhdgl 712
QY 109 VGM-----GIGAVASFVSSILFLFGSSKAKNSVAAYIDALSKHREAIQRNA 157
Db 713 ltlmalegfhvnglqghaslllaal-----kdgalvtvdagammkeesllqmaa 764
QY 158 G-AKRD-----PAESSAFIOVMKQOONLTDSDLSIIAANVPYKFS---NFIGOLESR1 207
Db 765 nqvekltkltswtqidaqlqwlqmsalavspidl--agmmalkygidhnyaaaw----- 817
QY 208 SGAATTSLSL-----AKRAVD--FLLYCOLVVMRETLVDLAILYRKGAHEHVASAVEN 261
Db 818 -qaaaaalmadhangqkklidelfskalcnyyl---navdsaaagvrdnnglytllidn 873
QY 262 -----ANRVKELADDTLDFLHKLIPEQALIGAVYHPISASETSKAILNTKRYGVPD 314
Db 874 qvsadvltstriaaiaaglylvnralnrdegqla-----sdvstrqfftdwery----- 922
QY 315 VPRPIGNRRKFTNSYVMNYISICSEA--YMGNYMFRGCSNVRNPNIRV--SKMSDGFYTM 370
Db 923 -----nkry-----stwgavselvyypeny-----dptgrlqgkmmadallqs 961
QY 371 EN-----SDRRKLYITKHDGKGWGTL-----DEDPGOGHMRFTPL----- 407
Db 962 lngsqlnadltvedafktyltstfeqvanlkvlsayhnhvndqgltyfigidqaapgytw 1021
QY 408 -----RHGKYVWSSKRMPNMFMYESSASGYIRSMEN 439
Db 1022 rsvdshkscengkf--aanaagew-----nkltcavnpwkn 1054

```

## RESULT 12

AAW17871  
ID AAW17871 standard; Protein; 2504 AA.

AC AAW17871;

DT 29-JAN-1998 (first entry)

DE Photorhabdus luminescens insect toxin protein TcdA.

XX Insecticide; insect; toxin; pest control; biological control;

XX Photorhabdus luminescens; TcdA; Southern corn rootworm;



CC toxic to insects upon exposure and especially when ingested. The  
 CC nucleic acid sequence can be used to produce transgenic plants.  
 CC baculoviruses or microbial hosts for toxin production. They can be used  
 CC to control insects pests from the lepidoptera, coleoptera, Hymenoptera,  
 CC diptera, Dictyoptera, Acarina or Homoptera orders, especially the  
 CC Southern or Western corn rootworm, Colorado potato beetle, mealworm,  
 CC boll weevil, turf grub, beetle armyworm, black cutworm, cabbage looper,  
 CC codling moth, corn earworm, European corn borer or tobacco hornworm  
 CC or budworm.  
 CC  
 CC  
 CC

SO Sequence 2504 AA.

Query Match 4.5%; Score 105.5; DB 19; Length 2504;  
 Best Local Similarity 19.2%; Pred. No. 11;  
 Matches 100; Conservative 85; Mismatches 184; Indels 151; Gaps 27;

QY 11 IYLAITSANHGKRS---DVNSLLTKVETALKEASGSNEALAL-----EG 53  
 DB 682 ltaclstlhgkeslgedlkrampcftsalhlt-sgevaqllllwidqigpaqltvdg 740  
 QY 54 LKEIOTKPRDVR---GQATKILGSVGSALGKLNAGDATKTIISGCLDIYAGIATTPGCP 108  
 DB 741 fweevqltptslkvlftfaqlaglslyrrig-lsetelslvtqgslllvagksllidhgi 799  
 QY 109 VGM-----GIGAVASFEVSSILSFTGSSAKNSVAAVIDRLSKHRDEAIORHAA 157  
 DB 800 ltlmalegfhtwnglqghaslllaal-----kdgalvtvdvagammkesllqmaa 851  
 QY 158 G-AKRD-----FAESSAFIQVKKQOQNLTDSDLSIIANVPYKFS---NFIGLESRI 207  
 DB 852 ngvekdltkltswtdalqlwlgmsalavspdl-agmalkygidhnyaa-----904  
 QY 208 SOGAATTSLSL---AKRAVD--FILLYQOLVVMRETLVLDALILRKGAHEHVASAVEN 261  
 DB 905 -qaaaaalnadhangaqkldetfskalcnyy1--navvdaaagyrdnglytllidh 960  
 QY 262 -----ANRVKELADPTLDFLHKLIPQALIGAVYHPIASSETSKALINTKRYGVPD 314  
 DB 961 gvsadvltstriaealagldlyvnrainrdgqla-----sdvstrqfttdewery-----1009  
 QY 315 VRRPIGNRRYKFTNSTYMTYSICSEA--YMGNYMERGCSNVRNPRIV--SKMSDGFYTM 370  
 DB 1010 -----nkry-----stwagvselyvypeny-----dpqtrigqtkmmdallqs 1048  
 QY 371 EN-----SDRRKLYITKHDGKGWGTL-----DEDPGDOGHMFITL-----407  
 DB 1049 lngsqlnadvedafkyltsfeyvanlkvisayhndvndgldlyfigldgaapqlyuw 1108  
 QY 408 -----RHGRYVSSKRMPNMFYMESSASGYIRSMEN 439  
 DB 1109 rsvdshkengkf--aanaagew-----nklcavnpwkn 1141

RESULT 14

AAB72610  
 ID AAB72610 standard; Protein: 2504 AA.

AC AAB72610;

XX 04-MAY-2001 (first entry)

DE Photorhabdus tcbd toxin.

XX Tcbda; Insect toxin; plant; insect resistance.

OS Photorhabdus sp.

XX WO200111029-A1.

XX 15-FEB-2001.

PF 11-AUG-2000; 2000WO-US22237.

XX 11-AUG-1999; 99US-0148356.  
 PR (DOWC ) DOW AGROSCIENCES LLC.  
 XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
 PI Sukhaphinda K, Merlo AO;  
 PI  
 DR WPI; 2001-191536/19.  
 DR N-PSDB; AAF58779.  
 PT Novel polynucleotide sequence encoding insect toxins, useful for  
 PT producing transgenic plants having resistance to insects, especially  
 PT corn rootworm  
 PS Disclosure; Page 62-72; 106pp; English.

The present invention provides the protein and coding sequences of modified versions of the Photorhabdus Tcbd and Tcbd toxins. These are suitable for expression in plants. The toxins are effective against insects upon ingestion, and the sequences provided can be used to produce transgenic plants with insect resistance. The present sequence is the Photorhabdus Tcbd protein.

SO Sequence 2504 AA:

Query Match 4.5%; Score 105.5; DB 22; Length 2504;  
 Best Local Similarity 19.2%; Pred. No. 11;  
 Matches 100; Conservative 85; Mismatches 184; Indels 151; Gaps 27;

QY 11 IYLAITSANHGKRS---DVNSLLTKVETALKEASGSNEALAL-----EG 53  
 DB 682 ltaclstlhgkeslgedlkrampcftsalhlt-sgevaqllllwidqigpaqltvdg 740  
 QY 54 LKEIOTKPRDVR---GQATKILGSVGSALGKLNAGDATKTIISGCLDIYAGIATTPGCP 108  
 DB 741 fweevqltptslkvlftfaqlaglslyrrig-lsetelslvtqgslllvagksllidhgi 799  
 QY 109 VGM-----GIGAVASFEVSSILSFTGSSAKNSVAAVIDRLSKHRDEAIORHAA 157  
 DB 800 ltlmalegfhtwnglqghaslllaal-----kdgalvtvdvagammkesllqmaa 851  
 QY 158 G-AKRD-----FAESSAFIQVKKQOQNLTDSDLSIIANVPYKFS---NFIGLESRI 207  
 DB 852 ngvekdltkltswtdalqlwlgmsalavspdl-agmalkygidhnyaa-----904  
 QY 208 SOGAATTSLSL---AKRAVD--FILLYQOLVVMRETLVLDALILRKGAHEHVASAVEN 261  
 DB 905 -qaaaaalnadhangaqkldetfskalcnyy1--navvdaaagyrdnglytllidh 960  
 QY 262 -----ANRVKELADPTLDFLHKLIPQALIGAVYHPIASSETSKALINTKRYGVPD 314  
 DB 961 gvsadvltstriaealagldlyvnrainrdgqla-----sdvstrqfttdewery-----1009  
 QY 315 VRRPIGNRRYKFTNSTYMTYSICSEA--YMGNYMERGCSNVRNPRIV--SKMSDGFYTM 370  
 DB 1010 -----nkry-----stwagvselyvypeny-----dpqtrigqtkmmdallqs 1048  
 QY 371 EN-----SDRRKLYITKHDGKGWGTL-----DEDPGDOGHMFITL-----407  
 DB 1049 lngsqlnadvedafkyltsfeyvanlkvisayhndvndgldlyfigldgaapqlyuw 1108  
 QY 408 -----RHGRYVSSKRMPNMFYMESSASGYIRSMEN 439  
 DB 1109 rsvdshkengkf--aanaagew-----nklcavnpwkn 1141

RESULT 15

AAB72612  
 ID AAB72612 standard; Protein: 2505 AA.

AC AAB72612;



```

XX 04-MAY-2001 (first entry)
DE Modified Photorhabdus tcbd toxin.
KM Tcbd; Insect toxin; plant; insect resistance.
OS Photorhabdus sp.
OS Synthetic.
PM WO200111029-A1.
XX 15-FEB-2001.
XX 11-AUG-2000; 2000MO-US22237.
PR 11-AUG-1999; 99US-0148356.
XX (DMC ) DOW AGROSCIENCES LLC.
PA
XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW,
PI Sukhapinda K, Merlo AO;
XX WPI; 2001-191536/19.
DR N-PSDB; AAF58781.
XX
PT Novel polynucleotide sequence encoding insect toxins, useful for
PT producing transgenic plants having resistance to insects, especially
PT corn rootworm.
XX
XX Claim 1; Page 83-93; 106pp; English.
XX
CC The present invention provides the protein and coding sequences of
CC modified versions of the Photorhabdus Tcbd and tcbd toxins. These are
CC suitable for expression in plants. The toxins are effective against
CC insects upon ingestion, and the sequences provided can be used to produce
CC transgenic plants with insect resistance. The present sequence is the
CC modified Photorhabdus Tcbd protein.
XX
XX Sequence 2505 AA;
XX
Query Match 4.5%; Score 105.5; DB 22; Length 2505;
Best Local Similarity 19.2%; Pred. No. 11;
Matches 100; Conservative 85; Mismatches 184; Indels 151; Gaps 27;
QY 11 IVLATSKKHGRS---DVNSLTKVERALKESGNSNALEAL-----BG 53
DB 683 Itatlstlghkeslgedikramapcftsalhlt-sgeyaydlilwidqipagltvdg 741
QY 54 LKGEIQTKPDYR-----GQATKILGVSALGKLSNGDPTKLIISGCLDIVAGIATTFGPG 108
DB 742 fweevqtptskvltfagvlslyrrig-lsetelsilvtqslivagksildhgl 800
QY 109 VGM-----GIGAVASVFSSILSFTGSSAKNSVAVIDRALSKHRDEAIQRHAA 157
DB 801 Itlmalegfhwvnglqghasillal-----kdgalvtvdvqamnkkesillgmaa 852
QY 158 G-AKRD-----FAESSAFIQVMKOOSNTBDSLIIAANVPYKFS---NFIGOLESKI 207
DB 853 ngvekdltkltswtqldallqwlqmsalavspidl--agmmalkygidhnyaa----- 905
QY 208 SOGAFTTSLSD---AKRAVD---FILLVQOLVVMRETLVDLAILYRKGNAEHVASAVEN 261
DB 906 -qaaataalmadhangakldetfskalcnnyl---navdsaaagvrdnnglylildn 961
QY 262 -----ANRVNKELAADTLDLFLKLIPEQALIGAVHPISASETSKALINTKYFGVDP 314
DB 962 qysadvltsrlaeataglqlynralnrdegla-----sdvstrqfftdwery----- 1010
QY 315 VPRPIGNRRYKFTNSYVNTYSCSEA--YMGNYMERGCSNVNPNIRV--SKMSDGFYTM 370
DB 1011 -----nkry-----stwagvselvyypeny-----dptqrlgqtkmmdallqs 1049

```

```

QY 371 EN-----SDRRKLYTKHDGWMGNTL-----DEDPDQGHMRFIPL----- 407
DB 1050 lngsqlnadtyedatfkylltsfegvanlkvisayhdnvnvqgltyflgldgaapgyyw 1109
QY 408 -----RHGRYVYSSKRWPMFMVMESSASGYIRSMEN 439
DB 1110 rsvdhskcengkf--aanaagew-----nklcavnpmkh 1142

```

Search completed: May 8, 2002, 12:13:09  
Job time: 122 sec